Uuk- H01590 91 P04215

The American Economic Review

ARTICLES

391
:001
V. 76
NO. 5
1986
91

J. YINGER **Measuring Racial Discrimination with Fair Housing Audits**

J. H. KAGEL AND D. LEVIN **The Winner's Curse and Public Information in Common Value Auctions**

P. REY AND J. TIROLE **The Logic of Vertical Restraints**

J. FARRELL AND G. SALONER **Installed Base and Compatibility**

J. A. BRANDER AND T. R. LEWIS **Oligopoly and Financial Structure**

M. AOKI **Horizontal vs. Vertical Information Structure of the Firm**

A. B. JAFFE **Technological Opportunity and Spillovers of *R&D***

J. R. MARKUSEN **Explaining the Volume of Trade**

P. M. GARBER **Nominal Contracts in a Bimetallic Standard**

J. B. DE LONG AND L. H. SUMMERS **Is Increased Price Flexibility Stabilizing?**

D. CARD **Efficient Contracts with Costly Adjustment**

W. J. BAUMOL **Productivity Growth, Convergence, and Welfare**

R. H. GORDON **Taxation of Investment and Savings in a World Economy**

D. LAM **The Dynamics of Population Growth, Differential Fertility, and Inequality**

SHORTER PAPERS: G. Johnson and G. Solon; R. J. Ruffin and F. Rassekh; J. R. Melvin; E. Lichtenberg and D. Zilberman; J. M. Poterba and L. H. Summers; C. L. Schultze; R. D. Milbourne and D. F. Richards; R. Eisner and P. J. Pieper; J. R. Barth, G. Iden, and F. S. Russek; F. Modigliani and A. Sterling; R. C. Kormendi and P. Meguire; A. G. Blomqvist; M. R. Carter; R. K. Sah and J. E. Stiglitz; E. Alvi; S. Bowles; W. G. Shepherd; M. Smirlock, T. Gilligan, and W. Marshall; E. K. Browning; J. A. Pechman; R. A. Margo; J. P. Smith.

DECEMBER 1986

THE AMERICAN ECONOMIC ASSOCIATION

- Printed at Banta Company, Menasha, Wisconsin. The publication number is ISSN 0002-8282.

- *THE AMERICAN ECONOMIC REVIEW* including four quarterly numbers, the *Proceedings* of the annual meetings, the *Survey*, and Supplements, is published by the American Economic Association and is sent to all members five times a year: March; May; June; September; December.

Regular member dues (nonrefundable) for 1987, which include a subscription to both the *American Economic Review* and the *Journal of Economic Literature* are as follows:

$37.50 if annual income is $30,000
$38.50 if annual income is $30,000 or less;
$46.20 if annual income is above $30,000, but no more than $40,000;
$53.90 if annual income is above $40,000;

Nonmember subscriptions will be accepted only for both journals: Institutions (libraries, firms, etc.), $110 a year; individuals, $72.00. Single copies of either journal may be purchased from the Secretary's office, Nashville, Tennessee.

In countries other than the United States, add $12.00 to the annual rates above to cover extra postage.

- Correspondence relating to the *Survey*, advertising, permission to quote, business matters, subscriptions, membership and changes of address should be sent to the Secretary, C. Elton Hinshaw, 1313 21st Avenue So., Suite 809, Nashville, TN 37212-2786. Change of address must reach the Secretary at least six (6) weeks prior to the month of publication. The Association's publications are mailed second class.

- Second-class postage paid at Nashville, Tennessee and at additional mailing offices. Printed in U.S.A.

- Postmaster: Send address changes to *American Economic Review*, 1313 21st Avenue So., Suite 809, Nashville, TN 37212-2786.

Founded in 1885

Officers

President
ALICE M. RIVLIN
The Brookings Institution

President-Elect
GARY S. BECKER
University of Chicago

Vice Presidents
PETER A. DIAMOND
Massachusetts Institute of Technology
MANCUR OLSON, JR.
University of Maryland

Secretary
C. ELTON HINSHAW
Vanderbilt University

Treasurer
RENDIGS FELS
Vanderbilt University

P4215

Managing Editor of The American Economic Review
ORLEY ASHENFELTER
Princeton University

Managing Editor of The Journal of Economic Literature
JOHN PENCAVEL
Stanford University

Executive Committee

Elected Members of the Executive Committee
VICTOR R. FUCHS
Stanford University
JANET L. NORWOOD
Bureau of Labor Statistics
ALAN S. BLINDER
Princeton University
DANIEL L. McFADDEN
Massachusetts Institute of Technology
SHERWIN ROSEN
University of Chicago
THOMAS J. SARGENT
University of Minnesota

EX OFFICIO Members
CHARLES L. SCHULTZE
The Brookings Institution
CHARLES P. KINDLEBERGER
Massachusetts Institute of Technology

THE AMERICAN ECONOMIC REVIEW

December 1986

VOLUME 76, NUMBER 5

Managing Editor
ORLEY ASHENFELTER

Co-Editors
ROBERT H. HAVEMAN
JOHN G. RILEY
JOHN B. TAYLOR

Production Editor
WILMA ST. JOHN

Board of Editors
CLIVE BULL
MICHAEL R. DARBY
JACOB A. FRENKEL
CLAUDIA D. GOLDIN
PHILIP E. GRAVES
GEORGE E. JOHNSON
JOHN F. KENNAN
MERVYN A. KING
MEIR KOHN
PAUL KRUGMAN
BENNETT T. McCALLUM
EDGAR O. OLSEN
ALVIN E. ROTH
STEVEN SHAVELL
JOHN B. SHOVEN
SUSAN WOODWARD
LESLIE YOUNG

- Submit manuscripts (4 copies), no longer than 50 pages, double spaced, to: Orley Ashenfelter, Managing Editor, *AER*; 209 Nassau Street, Princeton, NJ 08542–4607.

- Submission fee: $25 for members; $50 for nonmembers. Style guides will be provided upon request.

- No responsibility for the views expressed by authors in this *Review* is assumed by the editors or the publishers, The American Economic Association.

- Copyright © American Economic Association 1986. All rights reserved.

Articles

Measuring Racial Discrimination with Fair Housing Audits: Caught in the Act *John Yinger* 881

The Winner's Curse and Public Information in Common Value Auctions *John H. Kagel and Dan Levin* 894

The Logic of Vertical Restraints *Patrick Rey and Jean Tirole* 921

Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation *Joseph Farrell and Garth Saloner* 940

Oligopoly and Financial Structure: The Limited Liability Effect *James A. Brander and Tracy R. Lewis* 956

Horizontal vs. Vertical Information Structure of the Firm *Masahiko Aoki* 971

Technological Opportunity and Spillovers of *R&D*: Evidence from Firms' Patents, Profits, and Market Value *Adam B. Jaffe* 984

Explaining the Volume of Trade: An Eclectic Approach *James R. Markusen* 1002

Nominal Contracts in a Bimetallic Standard *Peter M. Garber* 1012

Is Increased Price Flexibility Stabilizing? *J. Bradford De Long and Lawrence H. Summers* 1031

Efficient Contracts with Costly Adjustment: Short-Run Employment Determination for Airline Mechanics *David Card* 1045

Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show *William J. Baumol* 1072

Taxation of Investment and Savings in a World Economy: The Certainty Case *Roger H. Gordon* 1086

The Dynamics of Population Growth, Differential Fertility, and Inequality *David Lam* 1103

Shorter Papers

Estimates of the Direct Effects of Comparable Worth Policy *George Johnson and Gary Solon* 1117
The Role of Foreign Direct Investment in U.S. Capital Outcomes *Roy J. Ruffin and Farhad Rassekh* 1126
The Nonequivalence of Tariffs and Import Quotas *James R. Melvin* 1131
The Welfare Economics of Price Supports in U.S. Agriculture *Erik Lichtenberg and David Zilberman* 1135
The Persistence of Volatility and Stock Market Fluctuations *James M. Poterba and Lawrence H. Summers* 1142
The Cyclical Flexibility of Wages *Charles L. Schultze* 1152
A New View of the Federal Debt and Budget Deficits:
  Comment *Ross D. Milbourne and Daniel F. Richards* 1154
  Reply *Robert Eisner and Paul J. Pieper* 1156
Government Debt, Government Spending, and Private Sector Behavior:
  Comment *James R. Barth, George Iden, and Frank S. Russek* 1158
  Comment *Franco Modigliani and Arlie Sterling* 1168
  Reply *Roger C. Kormendi and Philip Meguire* 1180
The Economics of Price Scissors:
  Comment *Ake G. Blomqvist* 1188
  Comment *Michael R. Carter* 1192
  Reply *Raaj Kumar Sah and Joseph E. Stiglitz* 1195
The Production Process in a Competitive Economy:
  Comment *Eskander Alvi* 1200
  Reply *Samuel Bowles* 1203
Tobin's q and the Structure Performance Relationship:
  Comment *William G. Shepherd* 1205
  Reply *Michael Smirlock, Thomas Gilligan, and William Marshall* 1211
Pechman's Tax Incidence Study:
  A Note on the Data *Edgar K. Browning* 1214
  A Response *Joseph A. Pechman* 1219
Race and Human Capital:
  Comment *Robert A. Margo* 1221
  Reply *James P. Smith* 1225

Notes 1230

Titles of Dissertations 1238

The following Statement of Ownership, Management, and Circulation is provided in accordance with the requirements, as contained in 39 U.S. Code 3685. The *American Economic Review* is owned, managed, and published by the American Economic Association, a nonprofit educational organization, located at 1313 21st Avenue South, Nashville, Davidson County, Tennessee 37212–2786. The Managing Editor is Professor Orley Ashenfelter, *American Economic Review*, 209 Nassau Street, Princeton, NJ 08542–4607. During the preceding 12 months the average number of copies printed for each issue was 27,878; the average paid circulation, 24,179; the average free distribution, 393; the average number of copies distributed, 24,572. Corresponding figures for the last issue before filing: 27,500 total number copies printed; 24,761 total paid circulation; 410 free copies distributed; 25,171 total distribution.

.391
.001

Measuring Racial Discrimination with Fair Housing Audits: Caught in the Act

By JOHN YINGER*

A survey technique called a fair housing audit provides a direct measure of racial discrimination in housing and an opportunity to test hypotheses about its causes. Audits conducted in Boston in 1981 uncovered extensive discrimination; black housing seekers were told about 30 percent fewer available housing units than were whites. This paper finds that the primary cause of this discrimination is economic: Housing agents cater to the racial prejudice of current or potential white customers.

With the help of a survey technique called a fair housing audit, several recent studies have demonstrated that, despite the Civil Rights Act of 1968, racial and ethnic discrimination in housing continues to be widespread. This paper examines the discrimination uncovered by a large audit study conducted in Boston in 1981 and analyzes the econometric issues that arise in evaluating fair housing audit data. The audit technique also makes it possible to test hypotheses about the causes of discriminatory behavior by housing agents. This paper states and carries out such tests.

In a fair housing audit, an individual from the white majority and an individual from a minority group, who have been matched according to their family and economic characteristics, successively visit a landlord or real estate broker in search of housing. By comparing the treatment minority and majority auditors receive, a researcher can isolate discrimination, which is defined as systematically less favorable treatment for the minority auditors.

Fair housing audits were developed by fair housing organizations as an investigative tool. They were first used for research in the pioneering study by Ronald Wienk et al. (1979) and have since been used in four other large studies, all of which pay careful attention to sampling, survey writing, auditor training, and survey management.[1] This paper examines data from the audit study conducted by Judith Feins, Rachael Bratt, and Robert Hollister.

Fair housing audits are an important methodological advance. Previous research examined discrimination indirectly by looking for its impacts on urban residential structure, such as higher prices in largely black areas and more centralized patterns of black residence (see my paper, 1979). Fair housing audits, on the other hand, literally catch discriminators in the act of discriminating. As a result, audits provide direct measures of discrimination and an unprecedented opportunity to study the circumstances under which discrimination occurs.

I. Measuring the Level of Discrimination

Fair housing audits yield information on the treatment black (or any other minority) and white housing seekers receive from hous-

---

*Professor of Economics and Public Administration, The Maxwell School, Syracuse University, Syracuse, NY 13244. I thank Judith Feins, who provided the data; Wilhelmina Leigh, Gary Solon, two anonymous referees, and participants in a seminar at the Institute of Public Policy Studies, University of Michigan, for helpful comments; and Howard Bloom and Michael Stoto for valuable advice on an early version of the material on audit methodology.

[1] The studies, all reviewed in my paper (1984), are Jon Hakken (1979); Judith Feins, Rachael Bratt, and Robert Hollister (1981) (summarized in Feins and Bratt, 1983); Franklin James, Betty McCummings, and Eileen Tynan (1983); and William Holshouser (1984). For a detailed review of study design and management issues, see Wienk et al., Feins et al., or myself (1984).

ing agents. In the broadest sense, racial discrimination exists whenever black auditors receive less favorable treatment than their white teammates. Treatment has many dimensions, however; each auditor is told about a certain number of units, quoted a certain rent or asking price, given certain signals about credit checks and financing, and so on. Although discrimination can occur in all types of treatment, this paper follows the existing studies in focusing on discrimination in the provision of information about available housing. After all, blacks cannot gain access to housing unless they know it exists.[2]

I employ three measures of housing availability: the number of housing units offered to the auditor as serious possibilities; the number of units the auditor was invited to inspect; and the number of units the auditor actually inspected.[3] For each measure, I estimate the level of discrimination and test the associated hypothesis that discrimination exists.

A. *A Simple Model of Discriminatory Treatment*

A fair housing "audit" consists of two or more "visits" to a housing "agent," some visits by black auditors and some visits by their white teammates. An audit study consists of audits carried out for a random sample of the advertised housing units in a given housing market using a set of trained auditors. Let us describe audit data with two subscripts: a identifies the audit and i identifies the individual auditor. The number of visits or observations equals the number of audits, say N, multiplied by the average number of auditors per audit. In most studies, including the 1981 Boston study, each audit consists of two visits, one by a minority auditor and one by a majority auditor, so the number of visits is $2N$.

Now let A measure an action taken by a housing agent toward an auditor, such as the number of apartments the auditor is invited to inspect; let R be a binary variable for minority status, where R equals one for a black and zero for a white auditor; and let e stand for a random error. Then the relationship between A and R is

$$A_{ai} = \alpha + \beta R_i + e_{ai}. \tag{1}$$

The ordinary least squares (OLS) estimate of β, say b, is a measure of the level of discrimination. If a higher value of A implies more favorable treatment, then a negative value for b implies that agents discriminate against minority auditors in action A. The t-test for β is a test of the hypothesis that discrimination exists. Note that b is the difference in the mean treatment of blacks and whites and the OLS t-test is equivalent to a standard difference-of-means test.

B. *Eliminating Bias Through the Audit Study Design*

The first statistical question to ask about equation (1) is whether b is biased. Many variables other than R influence the treatment an auditor receives, so the question is whether b is biased because these variables are left out of the equation. According to the well-known theorem, left-out-variable bias in b equals the sum, across all left-out variables, of the true coefficient of the left-out variable multiplied by the partial correlation between that variable and R (see J. Johnston, 1972, p. 169). The audit design eliminates bias by identifying variables that are likely to have large true coefficients—that is, large effects on A—and insuring that the correlation between these variables and R equals zero.

---

[2] For evidence about discrimination in other types of treatment, see my paper (1984). Discrimination in credit checks or the provision of mortgages is difficult to study with audits, but other methods can be used. See Robert Schafer and Helen Ladd (1981).

[3] Auditors try to visit all units to which they are invited, but not all invitations are genuine and they do not always succeed. One kind of discrimination, racial steering, may exist even if blacks and whites are told about the same number of available units; blacks may learn about units in black areas and whites about units in white areas. The Boston data provide some evidence of steering: after controlling for differences in the total number of units seen, on average black auditors saw .31 apartments and .21 houses that were not shown to their white teammates. See my paper (1984).

Three categories of left-out variables might influence treatment: whether the auditor precedes or follows his or her teammate(s); characteristics of the audit, including conditions in the housing market at the time of the audit and characteristics of the housing agent; and characteristics of the auditor. The audit design eliminates bias from all three categories.

Teammates cannot visit an agent at the same time, and their order of visit might influence their relative treatment. For example, a rental agent might discourage the second auditor if her teammate looked like a good prospect. To insure that order of visit would not be correlated with R, and therefore would not be a source of bias in b, Feins et al. assigned order randomly.

Sending auditors to the same agency during a short time span insures that agency characteristics and housing market conditions are the same for teammates and therefore are not correlated with R. Teammates visit the same agency but may encounter different agents. This possibility is not a source of bias. If the assignment of an agent to a customer is a random process, or if it depends on customer characteristics other than minority status, then variation in treatment across agents is not correlated with R. If, on the other hand, agent assignment depends on the customer's minority status, then differences in teammates' treatment due to the assignment process constitute discrimination and appropriately influence b.

The treatment an auditor receives is also influenced by his or her age, income, and family characteristics. In an audit study, teammates are matched on these characteristics, so that they are not correlated with R and therefore are not a source of bias. All the auditors in a team are assigned income and family characteristics. A study manager controls matching by teaming auditors with the same indelible characteristics, such as sex and age, and by assigning similar income and family characteristics to teammates.

All auditor characteristics that might influence treatment cannot be identified and incorporated into the matching process. Teammates may receive different treatment, for example, because of different personal styles. Audit studies minimize the potential bias from hard-to-measure auditor characteristics by giving all auditors the same thorough training and by avoiding auditors whose personalities might induce unusual treatment by agents.[4]

C. *An Error-Components Model of Discriminatory Treatment*

The audit study design implies that the random error in equation (1) has three components. The first, λ_a, is associated with audit a; that is, it measures variation in treatment due to factors shared by the teammates who participate in a given audit, including shared individual characteristics, agency characteristics, and market conditions. By definition, factors that are shared by teammates are not correlated with R. The second component, μ_i, refers to auditor i; that is, it measures variation in treatment due to auditor characteristics other than race that are not shared by teammates. Auditor assignment, matching, training, and selection are designed to insure that auditor characteristics with large impacts on treatment are reflected in λ_a, not μ_i. What remains in μ_i is small, random variation in the treatment of individuals that is not correlated with R. The third component, ε_{ai}, measures the component of treatment that is unique to auditor i's participation in audit a. Sending auditors to the same agency at about the same time insures that agency characteristics and market conditions, are reflected in λ_a, ε_{ai}. Randomizing order insures that order, the effect of which is in ε_{ai}, is not correlated with R. In symbols,

$$A_{ai} = \alpha + \beta R_i + \lambda_a + \mu_i + \varepsilon_{ai}. \tag{2}$$

Because the error components are not corre-

---

[4] Auditor training and selection and the assignment of roles eliminate auditor characteristics as a source of bias but also limit the generality of the results; an audit study does not reveal the level of discrimination against minorities with characteristics outside the range observed in auditors.

TABLE 1—THE LEVEL OF DISCRIMINATION IN HOUSING AVAILABILITY

	Measure of Availability					
	Rental Audits			Sales Audits		
	Possibilities	Invited To See	Inspected	Possibilities	Invited To See	Inspected
Measures of Treatment (Number of Housing Units)						
White Average = Constant Term	2.077	1.959	1.308	2.195	2.119	1.322
Level of Discrimination Against Blacks $= -b$	0.594	0.712	0.433	0.576	0.542	0.314
Test Statistics for the Level of Discrimination						
Bivariate Regression[a]						
Standard Error	0.153	0.150	0.101	0.172	0.178	0.097
t-Statistic[b]	3.877	4.734	4.302	3.343	3.055	3.247
R-Squared	0.046	0.067	0.056	0.046	0.038	0.043
Paired Difference-of-Means Test[c]						
Standard Error	0.119	0.123	0.089	0.125	0.139	0.077
t-Statistic[b]	5.007	5.779	4.883	4.623	3.902	4.074
R-Squared[d]	0.715	0.688	0.634	0.753	0.708	0.699
Number of Audits	156	156	156	118	118	118

Source: Author's calculations based on the Feins et al. 1981 Boston data.

Notes: Possibilities = number of housing units discussed as serious possibilities; Invited to see = number of housing units invited to inspect; Inspected = number of housing units actually inspected.

[a] And standard difference-of-means test.

[b] A t-statistic above 3.291 occurs by chance only once in 2000 trials.

[c] And bivariate regression plus audit dummies or *GLS* regression.

[d] Only applies to regression with audit dummies.

lated with R, the *OLS* estimate of β, namely b, is unbiased.

Although *OLS* yields an unbiased b, the *OLS* standard error for b is biased upward and could lead researchers or enforcement officials to conclude that discrimination does not exist when in fact it does. This bias arises because of the error component λ_a, which implies that the error terms for black and white auditors are correlated.[5] Consider the case of two auditors per audit. Let e_{aB} and e_{aW} be the error terms in equation (1) for the black and white auditors, respectively. Now define $\rho = \text{Cov}(e_{aB}, e_{aW})/\text{Var}(e_{ai})$. Then *OLS* overstates the standard error of b by a factor of $1/(1-\rho)^{1/2}$.[6] If $\rho = 0.5$, for example, the *OLS* formula overstates s_b by 41 percent.

With two auditors per audit, this error in s_b can be eliminated with a paired difference-of-means test, which uses the same b

---

[5] In principle, μ_i and the error component for order of visit, which is in ε_{ia}, could also lead to biased standard errors. A *GLS* procedure to account for μ_i would be complex but would add little; because the study design keeps μ_i small, its effect on the standard errors is small (for evidence, see fn. 11). Auditor dummies cannot be used to correct for μ_i because they are perfectly collinear with R. And as shown in Section II, a dummy variable for order, which corrects for the order error component, is never significant.

[6] In a bivariate regression of Y on X, the variance of b is

$$Var(b) = E\left[\sum_j w_j^2 e_j^2 + 2\sum_{j \neq k} w_j e_j e_k\right],$$

where j and k stand for observation, $w_j = x_j/\Sigma x_k^2$, and $x_j = (X_j - \bar{X})$. See Johnston (p. 20). In my case, $X_j = R_j$, $E(e_j^2) = \sigma^2$, and if teammates' observations appear together, $E(e_j e_k) = \text{Cov}(e_B, e_W)$ for teammates ($j = \text{odd}$, $k = j+1$) and $E(e_j e_k) = 0$ otherwise. Furthermore, $\rho = \text{Cov}(e_W, e_B)/\sigma^2$ (see text), $\Sigma w_k^2 = 1/\Sigma x_k^2$ (see Johnston, p. 19), and $x_j = (R_j - \bar{R})$ equals 0.5 for blacks and -0.5 for whites. Thus, $\text{Var}(b) = \sigma^2\{[1/[(.25)(2N)] + \rho(2N)(.5)(-.5)/[(.25)(2N)]^2\} = 2\sigma^2(1-\rho)/N$. It follows that the usual *OLS* formula, which assumes $\rho = 0$, overstates s_b, which is the square root of $\text{Var}(b)$, by a factor of $1/(1-\rho)^{1/2}$.

as the standard difference-of-means test, but incorporates the covariance between e_{aB} and e_{aW} into the calculation of s_b.[7] Identical results can be obtained from two regression models. A "fixed effects" model includes a dummy variable for each audit; the coefficient of the dummy variable for audit a is an estimate of λ_a. A "random effects" model treats the correlation between teammates' errors as part of the error structure and derives a generalized least squares (*GLS*) procedure to account for it. This *GLS* procedure is employed in Section II and is described in the Appendix.[8]

D. *Results from the 1981 Boston Audits*

Table 1 presents my analysis of the audit data collected by Feins et al. in Boston in 1981.[9] The second row contains the estimated level of discrimination in three measures of housing availability in both the rental and sales markets. The level of discrimination is high in all six cases.[10] For example, discrimination in "invitations" in the rental audits is 0.712 apartments; on average, blacks were invited to inspect 1.247 apartments, whereas white were invited to inspect 0.712 apartments more (or 57 percent more) than blacks. Overall, blacks must search much harder than whites to find equivalent housing; a black renter would have to visit 8 housing agents to receive as many invitations to inspect as a white receives from 5 agents.[11]

The correct t-test, in the seventh row of Table 1, indicates that these estimates of discrimination are highly significant. A comparison of third and sixth rows reveals that the incorrect bivariate regression yields standard errors that are biased upward by as much as 38 percent—but that still indicate significance. In a different audit study, a failure to correct a bias of this magnitude might lead one to conclude that discrimination does not exist when in fact it does.

II. Testing Hypotheses About Discriminatory Behavior

Table 1 documents that discrimination exists, but provides no insight into its causes. Discrimination is, of course, illegal regardless of its causes, but an analysis of causes helps to determine why discrimination persists and to design more effective policies to eliminate it. Consequently, this section explains and tests hypotheses about the causes of discriminatory behavior by housing agents.

---

[7] In the paired test, the t-statistic is the difference in average treatment of blacks and whites divided by an estimate of s_b equal to the square root of $[\text{Var}(A_{aW})+\text{Var}(A_{aB})-2\,\text{Cov}(A_{aB},A_{aW})]/N$.

[8] With two visits per audit and no control variables, these three methods are algebraically equivalent. This equivalence can be demonstrated with the results in Yair Mundlak (1978). A fixed effects model estimates β on the basis of variation in the data within groups (in my case within audits), but ignores variation between groups. A random effects model estimates β on the basis of a weighted average of within- and between-group variation. Because of the audit design, however, there is no between-group variation in R (every audit involves a zero and a one) and the *GLS* estimate collapses to the *OLS* estimate with audit dummies. Estimating each audit error is the same as analyzing differences between teammates, so all three methods are the same. Because a *GLS* procedure is consistent and asymptotically efficient, the other procedures share these desirable properties, and one might as well use the simplest procedure, namely the paired difference-of-means test. A more formal proof of these results is available from the author. I am grateful to Gary Solon for suggesting this line of argument.

[9] Feins et al. also present bivariate regression results for these three treatment variables. My results differ slightly from theirs because they weighted the data to count each neighborhood equally and they did not eliminate the bias in s_b.

[10] Audits avoid bias in b by insuring that teammates share factors that might influence treatment. Thus, a simple measure of the success of the design is the proportion of the variance in treatment explained by audit dummies. In Table 1, adding audit dummies boosts the R-squared from about 0.05 to about 0.7; most of the factors that influence treatment are indeed shared by teammates.

[11] Paul Courant (1978) shows that discrimination of this form can lead to a high degree of racial residential segregation and to a higher price for equivalent housing in black than in white areas. Schafer (1979) found large price differentials (up to 34 percent for some housing types) between largely black and largely white areas of Boston in 1970. The audit results and Courant's analysis suggest that these price differentials may have continued into the 1980's.

A. *The Hypotheses*

The literature on discrimination indicates that a housing agent might deny blacks access to housing because of his (or her) own prejudice or because of the prejudice of his white customers (see my paper, 1979). A third possibility is that housing agents treat blacks and whites differently because they expect blacks and whites to prefer different neighborhoods.[12] In this section, I apply these hypotheses to discrimination in the provision of information about available housing and test them using audit evidence. A similar set of hypotheses and tests (derived independently) for a different set of treatment variables is presented by John Simonson and Wienk (1984), using the audits carried out by Wienk et al.

The first hypothesis is that an agent's own prejudice leads him to discriminate in order to avoid dealing with blacks. According to this agent-prejudice hypothesis, variation in the strength of racial prejudice among housing agents leads to variation in the level of discrimination. Prejudice tends to increase with age, for example, and men tend to be more prejudiced than women (see George Simpson and Milton Yinger, 1985). More discrimination by older or by male brokers would therefore provide support for this hypothesis. The strength of an agent's prejudice, and hence the level of discrimination, also may vary with characteristics of the black auditor that matter to the agent but not to his potential white customers. The sex of a married black customer, for example, cannot matter to white residents (who would have both husband and wife as neighbors), but it might matter to the agent (who only deals with one spouse).

The second hypothesis is that an agent who depends on the business of a prejudiced white community discriminates in order to avoid alienating his current or potential white customers. Rental agents for all-white apartment buildings discriminate against blacks because the entry of a few blacks could lead to the exit of many or all white tenants and hence to extensive turnover costs. Real estate brokers cultivate contacts in their community to attract potential buyers and sellers. In a prejudiced white community, therefore, brokers discriminate against blacks to avoid alienating most of their potential clients, namely white residents.

These incentives to discriminate break down once a neighborhood starts to undergo racial transition. If blacks are willing to outbid whites for apartments, complete racial turnover may yield an increase in rental income that compensates a landlord for his turnover costs. And once racial transition begins, a real estate broker cannot be blamed for introducing blacks and he may be able to generate extra commission income by encouraging rapid turnover.

In sum, the white-customer-prejudice hypothesis predicts extensive racial discrimination in established white areas, but no discrimination in areas experiencing rapid racial transition.

The Boston audits took place in six neighborhoods, which are described in Table 2. Back Bay/Beacon Hill and Central are mostly white with significant increases in population but very small increases in black population between 1970 and 1980. The other four neighborhoods are more integrated, with racial compositions between 8 and 41 percent black, but they had widely varying patterns of racial change between 1970 and 1980. Dorchester and Hyde Park/Roslindale experienced large inflows of blacks and large outflows of whites—that is, they experienced rapid racial transition. Jamaica Plain lost white population, but gained few blacks. And the South End, which became a desirable location both for young, white professionals and for black families, saw rapid growth in both its black and white populations.

The customer-prejudice hypothesis predicts that discrimination will be high in Back Bay/Beacon Hill and Central and nonexistent in Dorchester and Hyde Park/Roslindale. It also predicts continuing discrimination in Jamaica Plain, where blacks' demand

---

[12] This hypothesis was proposed by an anonymous referee.

TABLE 2—CHARACTERISTICS OF STUDY NEIGHBORHOODS

	1980 Characteristics			1970–80 Change as Percentage of 1980 Population		No. of Sub-areas	Percentage of Owner Units in Subareas		Percentage of Renter Units in Subareas	
Neighborhood	Population	Percent Renter	Percent Black	Black	Other		< 20% Black	> 5% Black	< 20% Black	> 5% Black
BackBay/ Beacon Hill	30,212	82.1	4.2	2.2	6.7	1	100.0	0.0	100.0	0.0
Central	21,797	90.0	3.0	1.8	9.8	3	100.0	0.0	100.0	0.0
South End	27,125	86.9	40.8	7.7	8.7	1	0.0	100.0	0.0	100.0
Dorchester	82,912	65.2	24.7	11.3	−40.5	11	62.4	73.5	54.4	82.2
Hyde Park/ Roslindale	62,849	48.0	8.1	6.7	−25.3	9	75.3	38.4	82.6	36.0
Jamaica Plain	39,331	74.9	19.1	1.6	−23.1	5	72.5	59.5	44.4	85.4

Source: Boston Redevelopment Authority.

for housing does not appear to be growing, and in the South End, where many locations are currently attractive to whites.

White residents' prejudice, and hence agent discrimination, may also vary with the black auditor's characteristics. Survey evidence suggests that whites are more receptive to black neighbors with higher incomes and that whites may be more opposed to school integration than to residential integration, per se (see Simpson and Yinger). We can test the hypotheses that housing agents respond to these feelings by determining whether discrimination is stronger against either low-income blacks or blacks with school-age children.

The third hypothesis is that agents' actions reflect their perceptions about black and white preferences. According to surveys, most whites prefer not to move into an integrated neighborhood, and most blacks prefer an integrated neighborhood to an all-black one, although they also prefer not to be the first black in a previously all-white neighborhood (see Howard Schuman, Charlotte Steed, and Lawrence Bobo, 1985). Agents do not know the racial preferences of individual auditors; indeed, teammates are instructed to ask for the same housing and to express interest in all available units. On the other hand, agents may be aware of average preferences; this hypothesis assumes that agents attribute to each auditor the average preferences for his or her race. Agents' perceptions and actions are linked because showing apartments or houses is costly; an agent may not show a housing unit to a customer unless he believes there is a reasonable probability that the customer will rent or buy that unit.[13] According to the above assumption, therefore, an agent will not show housing in integrated neighborhoods to whites or housing in all-white neighborhoods to blacks.

To test this perceived-preference hypothesis, we must decide which neighborhoods are attractive to each race. One survey found that 50 percent of whites would "consider" moving into a neighborhood that was 21 percent black and that 95 percent of blacks would move into a neighborhood that was 14 percent black (Reynolds Farley, Suzanne Bianchi, and Diane Colasanto, 1979). If agents try to please the majority of their customers, it therefore seems reasonable to assume that any area less than 20 percent black is acceptable to whites, and any area greater than 5 percent black is acceptable to blacks. Several of the neighborhoods in the Boston study are large and heterogeneous, so the overall racial composition may not be an accurate guide to their attractiveness. Fortunately, a special census tabulation breaks the largest neighborhoods into relatively homogeneous subareas. Table 2 presents the percentage of owner and renter units in subareas that, according to the above assumptions, are acceptable to whites and to blacks.

---

[13] Simonson and Wienk identify several characteristics of black customers that might signal a low probability of completing a transaction—and therefore might lead to discriminatory treatment.

Based on the data in Table 2, the perceived-preference hypothesis predicts that discrimination will be relatively high in Back Bay/Beacon Hill and Central, where all the units are acceptable to whites and none are acceptable to blacks. In the South End, however, this hypothesis predicts no discrimination against blacks. (Indeed, the data in Table 2 indicate that no units are in areas acceptable to whites and therefore suggest that there might even be discrimination against whites. Subarea data are not available for the South End, however, so the heterogeneity mentioned earlier is not captured by these data.) In Dorchester and Jamaica Plain, the number of units acceptable to blacks and whites is about the same, so this hypothesis predicts no discrimination. In Hyde Park/Roslindale, many more units are acceptable to whites than to blacks, so this hypothesis predicts discrimination against blacks.

The perceived-preference hypothesis and the white-customer-prejudice hypothesis yield the same predictions in Back Bay/Beacon Hill, Central, and Dorchester. In the South End and Jamaica Plain, however, the customer-prejudice hypothesis predicts discrimination against blacks, whereas the perceived-preference hypothesis does not. The reverse is true for Hyde Park/Roslindale, where the perceived-preference hypothesis predicts discrimination, but the customer-prejudice hypothesis does not. We can decide which hypothesis to embrace by examining the levels of discrimination in these last three neighborhoods.

B. *Estimation Strategy*

Hypotheses about nondiscriminatory behavior are built into the audit study design. Teammates are matched on income, for example, because income is expected to influence their treatment. I now add direct controls for nondiscriminatory behavior by introducing into equation (2) the variables defined in Table 3. A variable to indicate the sex of the auditor, for example, controls for the possibility that female auditors are treated more favorably than males.[14]

To test the hypotheses about discriminatory behavior, let us introduce into equation (2) a series of interaction terms between race and the control variables in Table 3. Consider, for example, race interacted with whether the housing agent was female. A positive coefficient for this variable would indicate that female agents discriminate less than male agents. The interaction terms are expressed as deviations from a mean. For a particular auditor, the variable just mentioned, for example, equals the minority status of that auditor multiplied by the difference between the female-agent variable for that auditor and the average value of the female-agent variable for black auditors. This formulation allows us to continue to interpret the coefficient of the race variable as the average level of discrimination, as in equation (2). The coefficients of the interaction terms measure variation around this average.

Now let X^j_{ai} be control variable j for audit a and auditor i, and let γ_j be its coefficient. Furthermore, let $\bar{X}^j_{aB}$ be the average value of X^j_{ai} for all visits by black auditors. Then adding control variables and interaction terms to equation (2) yields

$$(3)\quad A_{ai} = \alpha' + \beta' R_i + \sum_j \gamma_j X^j_{ai} + \sum_j \delta_j R_i \left(X^j_{ai} - \bar{X}^j_{aB} \right) + \lambda'_a + \mu'_i + \varepsilon'_{ai}.$$

Applying *OLS* to equation (3) yields unbiased estimates of the parameters, but the controls do not fully eliminate the audit error component, λ'_a, and the standard errors of *OLS* coefficients are biased upward.[15]

---

[14] If matching eliminates correlation between the controls and R, then adding controls will not affect b. Table 4 shows that, by this test, matching was successful in the 1981 Boston audits.

[15] In the Boston audits, the R-squared with controls tends to be about half the R-squared with audit dummies, which fully capture λ_a; i.e., controls account for about half the variation in λ_a.

TABLE 3—DEFINITIONS AND AVERAGE VALUES FOR CONTROL VARIABLES

	Rental Audits		Sales Audits	
Variable	Black Auditor	White Auditor	Black Auditor	White Auditor
Auditor's Assigned Characteristics Income[a]	2.18	2.03	3.73	3.69
Role = single parent[b]			0.01	0.01
= childless couple	0.48	0.48	0.56	0.56
= couple with young children	0.04	0.04	0.25	0.27
= couple with older children			0.07	0.04
= previous owner			0.07	0.07
Auditor's Indelible Characteristics				
Age[c]	3.01	2.73	2.97	2.69
Sex (1 = Female)	0.54	0.54	0.47	0.47
Agent's Characteristics				
Race (1 = Nonwhite)	0.03	0.01	0.02	0.02
Young (1 = < 35-years-old)	0.57	0.57	0.36	0.36
Old (1 = > 50-years-old)	0.09	0.10	0.19	0.16
Sex (1 = Female)	0.49	0.46	0.46	0.48
Other Audit Information				
Auditor's order of visit (1 = 1st)	0.50	0.50	0.53	0.47
Audit was in June (not May)	0.46	0.46	0.33	0.35
Audit was in afternoon	0.57	0.61	0.58	0.64
Unit was a condominium[d]			0.37	0.37

[a] Income defined as follows: \$10,000 = 1; increments of \$10,000 to \$50,000 = 5; increments of \$15,000 to > \$80,000 = 7.

[b] Only one audit team was assigned the single-parent role, so the coefficient of this variable measures the relative treatment of this audit—not the relative treatment of single parents.

[c] Age defined as follows: < 25 = 1; 25–29 = 2; ...; 60–64 = 9; > 65 = 10.

[d] This variable counts audits in which either auditor was quoted a condominium fee, but misses condominiums if neither auditor was told that the advertised unit was available.

The *GLS* procedure in the Appendix can be used to eliminate this bias.[16]

C. *Results from the 1981 Boston Audits*

Tables 4 and 5 present the *GLS* results for variation in discriminatory behavior.[17] In Table 4, the coefficients of R, which still measure the average level of discrimination, are close in magnitude and significance to the estimates in Table 1. The other entries in Table 4 are coefficients of nonneighborhood interaction terms: the neighborhood results are in Table 5.

Few of the nonneighborhood variables have a statistically significant impact on dis-

---

[16] In equation (3), as in (2), X_a can be eliminated by differencing; i.e., by expressing the model as a difference between each black auditor and her white teammate. However, this formulation forces one to estimate the γ_js based solely on the within-audit variation in the X_js, despite the fact that most of the variation in these variables occurs between audits. This problem can be solved in two ways. First, unbiased estimates of the γ_js can be obtained using equation (3) and these estimates can then be used as a restriction in the differenced version of equation (3). This procedure has intuitive appeal, but unlike the *GLS* procedure in the text, its statistical properties are not known. (It turns out, however, that this paired approach and *GLS* yield similar results, at least for the 1981 Boston sample. Results for the paired approach are available from the author.) Second, one can ignore teammate differences in the X_js; after all, the audit design insures that these differences are small. This alternative is employed by Simonson and Wienk, but it is inappropriate for the Boston audits, because blacks are assigned slightly more favorable values than are their teammates for a few characteristics, such as income, and the left-out differences may be correlated with R. See Table 3.

[17] The *GLS* results for nondiscriminatory behavior are available from the author. Few controls are significant.

TABLE 4—ESTIMATED VARIATION IN DISCRIMINATORY BEHAVIOR

	Measure of Rental Availability			Measure of Sales Availability		
Variable	Possibilities	Invited To See	Inspected	Possibilities	Invited To See	Inspected
Average Level of Discrimination						
Auditor's Race	0.5757	0.6596	0.4502	0.4815	0.4972	0.3419
$(=R)$	(4.5510)	(5.0640)	(4.8490)	(3.1060)	(2.9470)	(3.8480)
Impact of Control Variables on the Level of Discrimination						
Auditor's Assigned Characteristics						
Income	0.1385	0.0849	0.0233	0.0657	0.0153	0.2068
	(0.7410)	(0.4390)	(0.1690)	(0.4960)	(0.1060)	(2.7380)
Childless Couple	−0.1305	−0.1710	−0.3554	0.7215	0.9129	−0.1863
	(0.4020)	(0.5080)	(1.4780)	(1.3600)	(1.5720)	(0.6150)
Couple, Child ≤ 6	−0.8675	−0.5022	−0.6397	0.1557	0.4617	−0.6757
	(1.1600)	(0.6490)	(1.1590)	(0.2630)	(0.7130)	(2.0030)
Couple, Child > 6				0.3417	−0.2387	−0.9226
				(0.3190)	(0.2040)	(1.5030)
Previous Owner				0.3648	0.8880	0.4768
				(0.4640)	(1.0350)	(1.0620)
Auditor's Indelible Characteristics						
Age	−0.109	0.0210	−0.0458	0.0621	0.0464	−0.1769
	(0.0400)	(0.0760)	(0.2330)	(0.1760)	(0.1220)	(0.8690)
Sex	0.2942	0.2818	0.2671	0.0976	0.2708	0.5896
	(0.6920)	(0.6530)	(0.8740)	(0.1880)	(0.4820)	(1.9710)
Agent's Characteristics						
Race	−1.0398	−1.1316	−1.2366	0.3100	0.1514	0.2948
	(0.8360)	(0.8640)	(1.3160)	(0.2850)	(0.1270)	(0.4760)
Young	0.3488	0.1641	0.0564	−0.2601	−0.1351	−0.0645
	(1.1390)	(0.5300)	(0.2580)	(0.7340)	(0.3540)	(0.3160)
Old	0.0953	−0.3136	−0.4013	0.0286	0.2139	0.2443
	(0.1890)	(0.6140)	(1.1090)	(0.0620)	(0.4340)	(0.9240)
Sex	−0.3339	−0.3684	0.2636	0.1701	−0.0855	0.1647
	(1.2410)	(1.3380)	(1.3460)	(0.5500)	(0.2550)	(0.9280)
Other Audit Information						
Order	0.1977	0.1884	−0.1139	0.1401	0.0876	−0.2430
	(0.5790)	(0.5760)	(0.5000)	(0.3640)	(0.2200)	(1.0650)
June	−0.2618	−0.2524	−0.1536	0.1961	0.2682	0.1175
	(1.0470)	(0.9740)	(0.8300)	(0.5660)	(0.7070)	(0.5940)
Afternoon	−0.5967	−0.6361	0.0612	−0.1503	0.2219	0.1052
	(2.0460)	(2.1570)	(0.2930)	(0.4430)	(0.6100)	(0.5360)
Condo				0.1396	0.3487	0.0357
				(0.3300)	(0.7540)	(0.1480)
R-Squared	0.2995	0.3172	0.2479	0.3647	0.3360	0.3317

Source: Author's calculations based on the Feins et al. 1981 Boston data.

Notes: Row 1 is $-b$ (-1 times the coefficient of R); the other rows are coefficients of variables in Table 3 interacted with R (a negative sign implies that discrimination increases with that variable). The sample is visits; dependent variables are defined in Table 1; *GLS* regressions also include control variables in Table 3, a dummy variable for each neighborhood in Table 2 (less 1), and neighborhood dummies interacted with R; interactions between R and the controls are expressed as deviations from the mean for blacks. Absolute values of t-statistics are shown in parentheses. The two-tailed 10 [5] (1) percent level of significance is 1.645 [1.960] (2.576).

criminatory behavior; in most cases, discrimination takes place regardless of the characteristics of the auditor or of the agent. In the rental market, discrimination in offering serious possibilities or in invitations to inspect is higher in the afternoon, after controlling for order—a result not predicted by any of the hypotheses. In the sales market, discrimination in actual inspections is stronger against low-income blacks and black families with children. (Estimated discrimination is even higher against black families with older children, although the coefficient is not quite significant at the 10 percent

TABLE 5—ESTIMATED LEVEL OF DISCRIMINATION BY NEIGHBORHOOD

	Rental Discrimination				Sales Discrimination			
Neighborhood	No. of Audits	Possibilities	Invited To See	Inspected	No. of Audits	Possibilities	Invited To See	Inspected
Sample Average	156	0.58[c]	0.66[c]	0.45[c]	118	0.48[c]	0.50[c]	0.34[c]
		(4.55)	(5.06)	(4.85)		(3.11)	(2.95)	(3.85)
BackBay/	46	0.68[c]	0.65[c]	0.07	16	0.03	0.04	0.75[c]
Beacon Hill		(2.97)	(2.74)	(0.43)		(0.59)	(0.07)	(2.57)
Central	31	0.97[c]	1.04[c]	0.98[c,d]	16	1.09[c]	0.99[b]	0.90[c,d]
		(3.40)	(3.49)	(4.61)		(2.44)	(2.04)	(3.54)
South End	36	0.05[d]	0.34[a]	0.42[b]	14	0.89[b]	0.61	0.63[c]
		(0.21)	(1.28)	(2.19)		(1.92)	(1.21)	(2.37)
Dorchester	11	0.17	0.55	0.61[b]	24	0.46	0.76[b]	−0.14[d]
		(0.34)	(1.06)	(1.65)		(1.18)	(1.79)	0.61
Hyde Park/	13	0.46	0.46	0.27	28	−0.06	0.11	−0.17[d]
Roslindale		(1.02	(1.00)	(0.80)		(0.16)	(0.29)	0.80
Jamaica Plain	19	0.99[c]	0.87[b]	0.60[b]	20	0.64[b]	0.61[a]	0.65[c]
		(2.80)	(2.37)	(2.30)		(1.70)	(1.48)	(3.02)

Notes: Sample average is from Table 4. Other entries equal (−1) times the coefficients of neighborhood dummies interacted with R from regressions like those described in Table 4, except that all six neighborhood interactions are included and R is excluded. Absolute values of t-statistics are shown in parentheses.

[a] = Significantly different from zero, 90 percent level.
[b] = Significantly different from zero, 95 percent level.
[c] = Significantly different from zero, 99 percent level.
[d] = Significantly different from average, 95 percent level.

level.) These results support the hypothesis that when it comes to actually showing houses to customers, real estate brokers are trying to keep out the black households that would most upset their potential white customers. In addition, black females encounter less discrimination in actual inspections than do black males; this result supports the hypothesis that some agents have particularly strong prejudice against black males and therefore avoid showing them houses.

Table 5 reveals substantial variation in discrimination across neighborhoods.[18] In Central and Jamaica Plain, the estimated level of discrimination in both the sales and rental markets is high and statistically significant for all three treatment variables. In Hyde Park/Roslindale, on the other hand, discrimination is low and insignificant and in one case is significantly below average. In Back Bay/Beacon Hill, the South End, and Dorchester, significant discrimination exists for one or two treatment variables, usually including actual inspections. For a black customer, housing agents may be willing to provide information, but they avoid taking the next step in a housing transaction, namely showing available units. The results for the sales market in Dorchester are somewhat contradictory; estimated discrimination in inspections is negative and significantly below the average level, despite significant discrimination in invitations to inspect.

These results provide strong support for the white-customer-preference hypothesis and a strong rejection of the perceived-preference hypothesis. As predicted by both hypotheses, there is discrimination in Back Bay/Beacon Hill and in Central. As predicted by neither, there is some discrimination in Dorchester, although in the sales market the discrimination in invitations appears to be canceled by equal treatment in actual inspections. In the other three neighborhoods, where the two hypotheses yield different predictions, the results all support the customer-preference hypothesis. To be specific, there is discrimination in the South End and in Jamaica Plain, but none in Hyde Park/Roslindale.

---

[18] In Table 5, the average level of discrimination is the coefficient of R in a regression with a set of neighborhood dummy variables (defined as deviations from the mean with one left out) interacted with R. The levels of discrimination by neighborhood are the coefficients of neighborhood dummy variables (defined as zero or one with none left out) interacted with R in a regression without R. These results measure discrimination by neighborhood because all auditors first asked about an advertised unit in a given neighborhood, and then asked about similar units in the same neighborhood.

Thus, housing agents discriminate primarily to retain the business of their actual or potential white customers. This incentive to discriminate breaks down when a neighborhood undergoes rapid racial transition; under these circumstances, discrimination may disappear (as in Hyde Park/Roslindale) or linger (as in Dorchester). Agent prejudice may be another source of discrimination in some cases, such as the nonshowing of houses to black males.

III. Conclusion

Well-managed fair housing audit studies provide direct, compelling tests of the existence of discrimination in housing. Unbiased measures of the level of discrimination (and its standard error) can be obtained with a very simple technique, namely the paired difference-of-means test. The application of this technique to the 1981 Boston audits reveals extremely high levels of discrimination. On average, for example, black auditors are invited to inspect 36.3 percent fewer apartments than their white teammates. Discrimination of this magnitude constitutes a serious social problem and cries out for stronger government action.

Fair housing audits also provide an unprecedented opportunity to study discriminatory behavior. In the Boston audits, racial discrimination in housing generally does not vary with the characteristics of the black auditor or the characteristics of the housing agent, although in actually showing houses for sale, it is stronger against lower-income blacks and black families with children. On the other hand, discrimination does vary by location, with high discrimination in white areas and no discrimination in some areas undergoing racial transition. These results indicate that the primary cause of racial discrimination in housing is that housing agents illegally promote their economic interests by catering to the racial prejudice of their current or potential white customers.

APPENDIX

To derive the *GLS* procedure, rewrite equation (1) in matrix form:

$$\text{(A1)} \qquad \mathbf{A} = \mathbf{R}\boldsymbol{\beta} + \mathbf{e},$$

where $E(\mathbf{e}) = 0$, $E(\mathbf{ee}') = \sigma^2\Omega$, and Ω is a symmetric, positive definite matrix. Now place visits for teammates next to each other. Then Ω has the following 2×2 blocks on the diagonal, zeroes elsewhere;

$$\text{(A2)} \qquad Z = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The *GLS* estimator is based on a matrix Q such that $Q'Q = \Omega^{-1}$. With audits, Q is a block diagonal matrix with Z replaced by

$$\text{(A3)} \qquad Z^* = \begin{bmatrix} w_1 & w_2 \\ w_2 & w_1 \end{bmatrix}$$

where

$$w_1 = \frac{1}{2}\left[\frac{1}{\sqrt{1+\rho}} + \frac{1}{\sqrt{1-\rho}}\right]$$

$$w_2 = \frac{1}{2}\left[\frac{1}{\sqrt{1+\rho}} - \frac{1}{\sqrt{1-\rho}}\right]$$

Equation (A3) defines a special case of Marc Nerlove's (1971) estimator.

To employ (A3), one must estimate ρ from the residuals of an *OLS* regression. With an estimated ρ, *GLS* is consistent and asymptotically efficient (see Thomas Fomby, R. Carter Hill, and Stanley Johnson, 1984, ch. 8), but second-stage observations for teammates are not independent and half the degrees of freedom are consumed. Second, one must premultiply **A** and **R** by Q. This matrix multiplication replaces an auditor's value for each variable, V, with a weighted average of that auditor's value and her teammate's value:

$$\text{(A4)} \qquad V_{ai}^* = w_1 V_{ai} + w_2 V_{ai'},$$

where i' denotes the teammates' visit. Applying *OLS* to the transformed variables yields the *GLS* estimator (Johnston, ch. 7).

REFERENCES

Courant, Paul N., "Racial Prejudice in a Search Model of the Urban Housing Market," *Journal of Urban Economics*, July 1978, *5*, 329–45.

Farley, Reynolds, Bianchi, Suzanne and Colasantro, Diane, "Barriers to the Racial Integration of Neighborhoods: The Detroit Case," *Annals of the American Academy of Political and Social Science*, January 1979, *441*, 97–113.

Feins, Judith D. and Bratt, Rachael G., "Barred in Boston: Racial Discrimination in Housing," *APA Journal*, Summer 1983, 344–55.

______, ______, and Hollister, Robert, "Final Report of a Study of Racial Discrimination in the Boston Housing Market," Abt Associates, November 1981.

Fomby, Thomas B., Hill, R. Carter and Johnson, Stanley R., *Advanced Econometric Methods*. New York: Springer-Verlag, 1984.

Hakken, Jon, *Discrimination Against Chicanos in the Dallas Rental Housing Market: An Experimental Extension of the Housing Market Practices Survey*, Washington: U.S. Department of Housing and Urban Development, 1979.

Holshouser, William, "Final Report of a Study of Racial Discrimination in Two Boston Housing Markets," Abt Associates, 1984.

James, Franklin J., McCummings, Betty L. and Tynan, Eileen A., "Discrimination, Segregation, and Minority Housing Conditions in Sunbelt Cities: A Study of Denver, Houston, and Phoenix," Center for Public-Private Sector Cooperation, Graduate School of Public Affairs, University of Colorado-Denver, 1983.

Johnston, J. *Econometric Methods*, 2nd ed., New York: McGraw-Hill, 1972.

Mundlak, Yair, "On the Pooling of Time Series and Cross Section Data," *Econometrica*, January 1978, *46*, 69–85.

Nerlove, Marc, "A Note on Error Components Models," *Econometrica*, March 1971, *39*, 383–96.

Schafer, Robert, "Racial Discrimination in the Boston Housing Market," *Journal of Urban Economics*, April 1979, *6*, 176–96.

______ and Ladd, Helen F., *Discrimination in Mortgage Lending*, Cambridge: MIT Press, 1981.

Schuman, Howard, Steeh, Charlotte and Bobo, Lawrence, *Racial Attitudes in America*, Cambridge: Harvard University Press, 1985.

Simonson, John C. and Wienk, Ronald E., "Racial Discrimination in Housing Sales: An Empirical Test of Alternative Models of Broker Behavior," unpublished manuscript, 1984.

Simpson, George E. and Yinger, J. Milton, *Racial and Cultural Minorities: An Analysis of Prejudice and Discrimination*, 5th ed., New York: Plenum Press, 1985.

Wienk, Ronald E. et al., *Measuring Discrimination in American Housing Markets: The Housing Market Practices Survey*, Washington: U.S. Department of Housing and Urban Development, 1979.

Yinger, John, "Prejudice and Discrimination in the Urban Housing Market," in P. Mieszkowski and M. Straszheim, eds., *Current Issues in Urban Economics*, Baltimore: Johns Hopkins University Press, 1979.

______, "Measuring Racial and Ethnic Discrimination with Fair Housing Audits: A Review of Existing Evidence and Research Methodology," paper submitted to the HUD Conference on Fair Housing Testing, Washington, D.C., December 6–7, 1984.

Boston Redevelopment Authority, *Boston Population and Housing by Neighborhood Area, 1980* (Demographic Information from the U.S. Bureau of the Census) September 1983.

The Winner's Curse and Public Information in Common Value Auctions

By JOHN H. KAGEL AND DAN LEVIN*

Experienced bidders show sensitivity to the strategic considerations underlying common value auctions, but not to item valuation considerations. Auctions with large numbers of bidders (6–7) produce more aggressive bidding than with small numbers (3–4), resulting in negative profits, the winner's curse. Providing public information about the value of the item increases seller revenue in the absence of a winner's curse, but produces the contrary result in its presence.

Common value auctions constitute a market setting in which participants may be particularly susceptible to judgment failures that affect market outcomes. In a common value auction, the value of the auctioned item is the same to all bidders. What makes the auction interesting is that bidders are unaware of the value of the item at the time the bids are placed. Mineral lease auctions, particularly the federal government's outer continental shelf (OCS) oil lease auctions, are common value auctions. There is a common value element to most auctions. Bidders for an oil painting may purchase for their own pleasure, a private value element, but they may also bid for investment and eventual resale, reflecting an uncertain common value element.

Judgmental failures in common value auctions are known as the "winner's curse." Assume that all bidders obtain unbiased estimates of an item's value and that bids are an increasing function of these estimates. The high bidder then tends to be the one with the most optimistic estimate of the item's value. Unless this adverse-selection problem is accounted for in the bidding process, it will result in winning bids that produce below normal or even negative profits. The systematic failure to account for this adverse selection problem is referred to as the "winner's curse."

Oil companies claim they fell prey to the winner's curse in early OCS lease sales (E. C. Capen, R. V. Clapp, and W. M. Campbell, 1971; John Lorenz and E. L. Dougherty, 1983; and references cited therein). Similar claims have been made in auctions for book publication rights (John Dessauer, 1981) in professional baseball's free agency market (James Cassing and Richard Douglas, 1980) and in corporate takeover battles (Richard Roll, 1986). Economists typically treat such claims with caution as they imply that bidders repeatedly err, in violation of basic notions of economic rationality. This caution is justified given the inherent problems in interpreting field data, self-serving motives of many of the claimants, and the general absence of conventional statistical tests documenting these claims. However, common value auction experiments using financially motivated, but inexperienced, subjects demonstrate a strong and nearly ubiquitous winner's curse which continues to significantly depress profits after as many as 15–20 auction periods (Kagel et al., 1986).

---

*Department of Economics, University of Houston-University Park, Houston, TX 77004. Financial support was received from the Information Science and Technology Division and the Economics Division of NSF; the Energy Laboratory and the Center for Public Policy of the University of Houston-University Park. Ray Battalio, Don Meyer, and Carl Kogut ran auction series 1 and 2; Ron Harstad and Doug Dyer assisted in series 3–8. Doug Dyer and Susan Garvin provided valuable research assistance. This paper has benefited from discussions with Badi Baltagi, Ron Harstad, Mark Issac, Asbjord Moseidjord, Jim Smith, and Philip Sorensen, the comments of Daniel Friedman, Robert Wilson, two referees, and participants in seminars at the University of Houston, Texas A&M University and the University of Indiana. An earlier version of this paper was presented at the 1985 Winter Econometric Society meetings. We alone are responsible for errors and omissions.

It is one thing to find that inexperienced bidders commit a winner's curse. It is another to find that experienced bidders do the same. Here we report the results of common value auction experiments with experienced subjects, survivors of one or more initial series of experiments. With continued experience, bidders' judgment improve. In auctions involving a limited number of competitors (3–4 bidders), average profits are consistently positive and closer to the Nash equilibrium bidding outcome than to the winner's curse hypothesis: behavior consistent with traditional notions of the effects of repeated exposure to market conditions, in conjunction with profit incentives and survival pressures. However, learning is situationally specific as bids are found to be an increasing function of the number of rivals faced, in clear violation of risk-neutral Nash equilibrium bidding theory under our design. This contributes to a reemergence of the winner's curse, with bankruptcies and negative profits, in auctions with large numbers (6–7) of bidders.

Just as Nash equilibrium bidding theory predicts, experimental manipulations providing public information reducing uncertainty about item value reliably result in higher winning bids and increased seller's revenues in the absence of a winner's curse. However, in the presence of a winner's curse, this same public information generates lower average winning bids and reduced seller's revenues. The differential response to public information conditional on the presence or absence of a winner's curse has practical implications which have largely gone unrecognized in the literature.

The paper is organized as follows. Section I describes the structure of the experiments. Section II characterizes the Nash equilibrium bidding strategies for the auction, provides a formal definition of the winner's curse, and states the research hypotheses that guided our investigations. The results of the experiments are reported and discussed in Section III. Section IV extends the analysis to field settings where our experimental results help explain a puzzling outcome of OCS lease sales: namely, that public information reducing item uncertainty increased bidder's profits, just as observed in our laboratory experiments in the presence of a winner's curse. A concluding section summarizes our research results and poses questions for further research.

I. Structure of the Auctions

A. *Basic Auction Structure*

Subjects were recruited for two-hour sessions consisting of a series of auction periods. In each auction period, a single unit of a commodity was sold to the high bidder at the high-bid price, with bidders submitting sealed bids for the item (a first-price, sealed-bid procedure). The high bidder earned profits equal to the value of the item less the amount bid; other bidders earned zero profits for that auction period.

In each auction period, the value of the item, x_0, was drawn randomly from a uniform distribution on the interval $[\underline{x},\bar{x}]$. Subjects submitted bids without knowing the value of x_0. Private information signals, x_i, were distributed prior to bidding. The x_i were randomly drawn from a uniform distribution centered on x_0 with upper bound $x_0+\varepsilon$ and lower bound $x_0-\varepsilon$. As such, the x_i constitute unbiased estimates of the value of x_0 (or could be used to compute unbiased estimates in conjunction with the endpoint values $\underline{x},\bar{x})$. Given x_i, ε, and the endpoint values, each bidder could compute an upper and lower bound on the value of x_0; these were $\min\{x_i+\varepsilon,\bar{x}\}$ and $\max\{x_i-\varepsilon,\underline{x}\}$, respectively. The bounds associated with a given x_i were computed and reported along with the x_i.

The distribution underlying the signal values, the value of ε, and the interval $[\underline{x},\bar{x}]$ were common knowledge. The value of ε varied across auctions (see Table 1). All changes in ε were announced and posted. With signal values drawn independently relative to x_0, they satisfy the criteria of strict positive affiliation (Paul Milgrom and Robert Weber, 1982) which, roughly speak-

TABLE 1—EXPERIMENTAL CONDITIONS[a]

Auction Series	Subject Population (no. starting exp.)	Market Period	ε	$\underline{x}$	$\bar{x}$	Public Information[b] (market periods)	Number Active Bidders (market periods)	Experience
1	Texas A&M Undergraduates (5)	1–18	\$12	\$15	\$100	Random signal (10–18)	5(1–18)	First-price common value
2	Texas A&M Undergraduates (4)	1–18	\$12	\$15	\$100	Random signal (10–18)	4(1–18)	First-price common value
3	U. Houston Graduate/Senior Undergraduates (7)	1–8 9–17,24–25 18–23	\$12 \$18 \$30	\$25	\$225	Low signal (12–25)	7(1–4) 6(5–6) 5(7–11) 4(12–25)	Second-price common value, some first-price common value
4	U. Houston Graduate/Senior Undergraduates (8)	1–6 7–14 15–25	\$12 \$18 \$30	\$25	\$225	Low signal (8–25)	7(1–10) 6(11–25)	Second-price common value, some first-price common value
5	U. Houston Graduate/Senior Undergraduates (9)	1–9 10–15,24–26 16–23	\$12 \$18 \$30	\$25	\$225	Low signal (7–26)	7(1–26) 7(1–26)	First and Second-price common value
6	U. Houston Graduate/Senior Undergraduates (4)	1–5 6–16,28–31 17–21 22–27	\$12 \$18 \$24 \$30	\$25	\$225	Low signal (9–31)	4(1–11) 4(1–11) 3(12–31)	First and Second-price common value
7	U. Houston Graduate/Senior Undergraduates (6)	1–5 6–13,26–32 33–37 14–25	\$12 \$18 \$24 \$30	\$25	\$225	Low signal (9–32)	4(1–19) 6(20–37)	First and Second-price common value
8	U. Houston Graduate/Senior Undergraduates (7)	1–6 7–16 17–23	\$12 \$18 \$30	\$30	\$500	None	4(1–23)[c] 7(10–23)	First-price common value, some private value

[a] Starting balances were $8 in experiments 1 and 2, $10 in all others.

[b] Profits were earned in markets with both public and private information in experiments 1 and 2; in only one market in experiments 3–8. The market paying profits was determined by a coin flip in experiments 3–8.

[c] Period 10 on involved a bidding in two markets with 4 subjects bidding first in a "small" market and all 7 subjects bidding in a "large" market.

ing, requires that large values for a given signal make it more likely that rivals signal values, and x_0, are large rather than small.

Bids were restricted to be nonnegative and rounded to the nearest penny. After all bids were collected, they were posted on the blackboard in descending order next to the corresponding signal values, x_0 was announced, subjects' profits were calculated, and balances were updated.[1] Earnings of the high bidder were also announced, but his/her identity was not. The x_0 values and the associated signal values were all determined randomly strictly according to the process described to the subjects.

To cover the possibility of losses, subjects were given starting balances of $8.00 in auction series 1–2, and $10.00 in series 3–8. Profits and losses were added to this balance. If a subject's balance went negative, he was no longer permitted to bid; he was paid the $4.00 participation fee and free to leave the experiment. The auction survivors were paid their end of experiment balance in cash, along with their participation fee.[2]

---

[1] In auction series 1 and 2, the top three bids were posted, and the signal values underlying the bids were not revealed.

Given the information structure and uncertainty inherent in common value auctions, negative profits will occasionally be realized even if the market immediately locks into the risk-neutral Nash equilibrium outcome. The starting capital balance served to account for this possibility, and to impose clear opportunity costs on overly aggressive bidding. Balances were set so that: (*i*) subjects could commit at least one gross bidding error, learn from their mistake, and still have a large enough balance to actively participate in the auctions, and (*ii*) conservative bidders who were shut out from winning by overly aggressive counterparts would earn a reasonable return for participating.

B. *Auctions with Public Information*

Approximately one-third of the way through auction series 1–7, we introduced bidding in two separate auction markets simultaneously. Bidding in the first auction market continued as before under private information conditions. After these bids were collected, but before they were posted, we introduced a public information signal and asked subjects to bid again. (Subjects retained their original private information signals; no new private information signals were distributed.) We employed two types of public information signals. In series 1 and 2, we randomly drew an additional signal, x_i, from the interval $[x_0-\varepsilon, x_0+\varepsilon]$, and posted it. In series 3–7, the lowest of the private information signals distributed, x_L, was posted. Bidders were always accurately informed of whether the public information signal was random or the lowest private information signal.

Profits were paid (or losses incurred) in only one of the two auction markets, determined on the basis of a coin flip after all bids were collected.[3] Subjects were told that they were under no obligation to submit the same or different bids in the two markets, but should bid in a way they thought would "maximize profits." All bids from both markets were posted along with the corresponding private information signals.

The dual market bidding procedure, involving the same set of bidders with the same item value and the same set of private information signals, has the advantage of directly controlling for between subject variability and extraneous variability resulting from variations in item value and private information signals. Some critics have mistakenly concluded that the procedure involves a "portfolio" problem so that the optimal bid in one market affects bids in the other market. This conclusion is unwarranted, however. There is no way that bids in the private information market can be used to hedge bids in the public information market any more than bids in the private information market in period t can be used to hedge bids in the private information market in period $t+1$. In analyzing each member of a set of auctions, $t=1,2,\dots,T$, as a single-shot auction (which we do below), we are assuming that the utility function is intertemporally separable in profits from the auction, $U=\Sigma_{t=1}^{T}u(\Pi_t)$. Similarly we are assuming separability in bidding between the dual markets. The breakdown of the separability assumption in either case (for example, $U=u(\Sigma_{t=1}^{T}\Pi_t)$) has comparable implications, namely no effect in the case of risk-neutral bidders, while risk-averse bidders will tend to be less risk averse than under separable preferences, as they can rely on the law of large numbers to smooth the variance in profits across auctions.

Of course, it is another matter entirely whether bidders actually bid as if their preferences are separable between the dual markets or over time. Experiments 1 and 2 explicitly tested for separability, holding ε constant throughout and having a relatively large number of private information auctions (9) before introducing dual markets. Regression analysis using dummy variables showed no systematic response to bids in the private information market under the dual market procedure.[4] Similar tests, conducted in a related series of first-price auctions, showed no systematic effects in going from private information markets only to dual markets with public information, and from dual markets to markets with public information only (Kagel, Ronald Harstad, and Levin, 1986). Bid patterns over time are presented in reporting our results.

---

[2] Series 1 and 2 had a $3 participation fee. Many subjects in series 3–7 had signed a contract to participate in 3 different auction series, for which they were paid a single participation fee of $25 at the end of the last series.

[3] Once again, series 1 and 2 involved an exception to these procedures as profits and losses were computed and paid in both the public and private information markets.

C. *Varying Numbers of Bidders*

A number of tactics were employed to study the effects of varying the number of bidders. Auction series 1–6 used a between-groups design with different series having different numbers of active bidders. In pursuit of this objective, series 4 and 5 had more subjects than active bidders, in order to control for bankruptcies, with a simple rotation rule to determine which subjects would be active in any given auction period.[5] Variations in the number of active bidders in these series (see the eighth column in Table 1) resulted from bankruptcies.

Auction series 7 and 8 involved planned variations in the number of active bidders. Series 7 employed a crossover design, starting with 4 active bidders rotating among a set of 6 total bidders. In auction period 20, the rotation procedure ceased, and all 6 bidders were active in the remaining auction periods. Auction series 8 employed a within-subjects design, starting out with 4 active bidders rotating between 7 total bidders. In auction period 10, dual market bidding procedures similar to those used to study the effects of public information were introduced; only numbers of active bidders varied between markets.[6] We refer to the different market sizes here as series 7 and 8 small, and 7 and 8 large. Auction series 3 involved a large unplanned variation in number of active bidders due to bankruptcies in early auction periods. Our analysis distinguishes between these early auction periods with 5 or more active bidders (series 3 large), and the later periods with 4 active bidders (series 3 small), as market outcomes were distinctly different between the two situations.

D. *The Experience Factor*

All auctions employed experienced subjects. In series 1 and 2, all subjects had been in one earlier first-price common value auction series using similar design parameters (see Kagel et al., 1986, auction series 4–6). These earlier auctions all began with 6 active bidders, but as a result of bankruptcies ended with 3–5 bidders. Recruitment into these experiments was restricted to subjects who had not gone bankrupt in the initial auction series.

Auction series 3–8 are numbered in chronological order as they involved a common core of subjects, recruited in varying combinations, in the different series. Thirteen of the 15 bidders in series 3 and 4 had participated in an earlier series of second-price common value auction experiments with similar design parameters. (The distinguishing characteristic of these second-price auctions was that the high bidders earned the item and paid the second-highest bid price.) Most of these subjects had been in two or more second-price series, at least one of which involved 6–7 active bidders throughout. The two remaining subjects had been in a first-price common value series involving 5–6 active bidders throughout (Kagel et al., 1986, auction series 11). Seven of the 9 subjects in series 5 were recruited from series 3 and 4, with the remaining 2 having extensive second-price experience. All 6 bidders in series 7 had been in both auction series 3 and 5, or 4 and 5. Three of the 4 bidders in series 6 had been in series 4, with the fourth bidder having been in series 3 and 5. Series 8 was conducted several months after the others and involved 3 veterans of series 7, 2 from series 6, and 2 bidders with experience in first-price auctions with positively affiliated private values (Kagel, Harstad, and Levin). Several subjects in this series had gone bankrupt in their initial common value auction series, but were included provided the bankruptcy occurred after a fair number of auction periods (15 or more periods was the rule of thumb employed).[7] Subjects were recruited into later series in this sequence without regard to performance earlier in the sequence.

---

[4] The bid function had the form $b_{it} = \alpha_0 + \alpha_1 x_{it} + \alpha_2 Y_{it}$ as suggested by equation (2), where α_i were constants to be estimated (ε was constant in these auctions). The function was fit to individual subject data for signals in the interval (1), and a dummy variable added to account for simultaneous bidding in two auctions. Combining independent t-tests on the dummy variable coefficient using the z statistic suggested in Ben Winer (1971, p. 50), we were unable to identify any systematic effects associated with simultaneous bidding ($z = -.124$).

[5] For example, with 9 subjects and 7 active bidders, in period 1, subjects 1–7 were active; in period 2, subjects 2–8 were active, etc. Inactive subjects received signals and bid, but these bids were discarded (the latter was common knowledge).

[6] Using a single set of private information signals and a single true value, x_0, 4 subjects first bid in a small market. Then, before these bids were opened, all 7 subjects bid in a large market. The 4 subjects in the small market continued to be determined through rotation, and bids only counted in one of the two markets using a coin flip rule. Raymond Battalio, Carl Kogut, and Donald Meyer (1983) report tests of the separability assumption underlying the dual market technique in private value auctions with varying numbers of bidders. They found no systematic biases associated with the dual market technique.

II. Theoretical Considerations

A. *Private Information Conditions*

1. *The Nash Equilibrium.* The most common equilibrium bidding model in the economic's literature is that of a noncooperative Nash equilibrium with risk-neutral bidders (hereafter RNNE). Robert Wilson (1977) was the first to develop a Nash equilibrium solution for first-price common value auctions, while Milgrom and Weber provide some significant extensions and generalization of the Wilson model.

We restrict our analysis in the text to signals in the interval

$$(1) \qquad \underline{x}+\varepsilon \le x_i \le \bar{x}-\varepsilon.$$

The optimal bid function in this interval is, of course, affected by the bid function in the interval $x_i < \underline{x}+\varepsilon$, which in turn is affected by the added information associated with the end-point value, $\underline{x}$.[8] Assuming risk neutrality on bidder's part, the Nash equilibrium bid function for signals in (1) is

$$(2) \qquad b(x_i) = x_i - \varepsilon + Y,$$

where $Y = [2\varepsilon/(N+1)]\exp[-(N/2\varepsilon)(x_i - (\underline{x}+\varepsilon))]$ and N stands for the number of active bidders in the market. Y contains a negative exponential, and diminishes rapidly as x_i moves beyond $\underline{x}+\varepsilon$.

Under (2), expected profits for the high bidder are

$$(3) \qquad E[\Pi|W] = 2\varepsilon/(N+1) - Y.$$

In addition, the model predicts that the high signal holder always wins the auction. This follows directly from the assumption that all bidders use the same bid function, the only

---

[7] In a private value auction, bidders know their value for the item with certainty, but only the distribution of their rivals' values. Under the first-price rule, the high bidder wins the item and earns profits equal to his private value less the bid price; others earn zero profits.

[8] For $x_i < \underline{x}+\varepsilon$, the RNNE bid function is

$$b(x_i) = \underline{x} + (x_i + \varepsilon - \underline{x})/(N+1)$$

and yields zero expected profits. This equilibrium bid function is obtained from Wilson (1977) under the initial condition $b(x_i) = \underline{x}$ for $x_i = \underline{x} - \varepsilon$. The initial condition for the bid function in (2) exploits continuity in the bid function at the junction point $x_i = \underline{x}+\varepsilon$. For $x_i > \bar{x}-\varepsilon$, the optimal bid function defies analytic solution. For observations in this interval, we employ the bid function (2) in comparing performance with the RNNE model. This tends to overstate the RNNE bid, hence underestimate the discrepancy between actual and predicted bids. Since the bias is small, and favors the null hypothesis, correcting for it will not change the conclusions reached. These bid functions are explicitly derived in Kagel et al. (1984).

P4215

difference being their private information, x_i, regarding the value of the item.[9]

Accounting for risk aversion on bidder's part complicates the model's predictions, as equilibrium bids can lie to either side of the RNNE prediction, depending upon the form of the utility function and the degree of risk aversion assumed. We do not pursue these extensions here, as they appear secondary to understanding the experiments' outcomes.[10] Note that under the RNNE there is no winner's curse as bidders fully account for the adverse-selection problem in determining their bids. The RNNE bidding model provides a convenient benchmark against which to compare the experiments' outcomes.

2. *Judgmental Failures: The Winner's Curse.* In common value auctions, bidders usually win the item when they have the highest, or one of the highest, estimates of value. Under these conditions an unbiased estimate of value, $E[x_0|x_i)$, is biased as

$$E[x_0|x_i] > E[x_0|X_i = x_1] \qquad \text{for } N > 1$$

where $E[x_0|X_i = x_1]$ is the expected value conditional on having the highest private information signal. Assuming that the highest signal holder always wins the auction and risk neutrality, or risk aversion, bids in excess of $E[x_0|X_i = x_1]$ will insure negative profits on average, and can only result from failure to recognize the adverse-selection problem inherent in winning the auction. Since we have no reason to assume risk loving, and the highest signal holder usually wins the auction, bids in excess of $E[x_0|X_i = x_1]$ will be attributed to such judgmental failures, and will be referred to as the winner's curse.[11]

For signal values in the interval (1),

$$(4) \quad E[x_0|x_i] = x_i;$$

$$(5) \quad E[x_0|X_i = x_1] = x_i - \varepsilon(N-1)/(N+1).$$

Avoiding the winner's curse requires considerable discounting of bids relative to signal values.[12] Further, the size of the discount is an increasing function of both N and ε.

In first-price sealed-bid auctions, strategic considerations generally dictate discounting of bids relative to the expected value of the item. Strategic discounting results strictly from known dispersion in rivals' values. It is informative to compare the bid function (2), and the size of the discount in (5), with strategic discounting based on the dispersion in the x_i values. Suppose that bidders completely ignore the adverse-selection problem inherent in the auction, employing (4) to compute the expected value of the item: they act as if they are in an auction with positively affiliated private values, where the x_i represents the value of the item to bidder i, and values are independently distributed over the interval (1). Under risk neutrality, the bid function here is[13]

$$(6) \quad b^s(x_i) = x_i - (2\varepsilon/N) + (Y/N),$$

---

[9] The model can be generalized to account for both private value and common value elements (see Wilson, 1981, for example). In this case, the high signal holder does not always win the item, even assuming identical bidding strategies. Undoubtedly, actual auction sales, even OCS lease sales, contain both private and common value elements. Developments in Section III suggest that the Nash equilibrium model must be generalized to account for individual differences in risk attitudes and/or information-processing capacities, as well as simple random errors on bidder's part. Developing and testing such a model lies beyond the scope of the present paper.

[10] Milgrom and Weber, section 8, and Steven Matthews (1986) develop some results for risk-averse bidders.

[11] Results from private value auction experiments generally support an assumption of risk neutrality or risk aversion (James Cox, Bruce Roberson, and Vernon Smith, 1982; Kagel, Harstad, and Levin).

[12] The size of the discount seems remarkably large, particularly for larger N's. To get an idea of the adverse-selection bias involved here, compared to alternative distributions, we normalize it in terms of the standard error of the underlying distribution. For x_i in (1) the standard error of the uniform distributions is $\sigma = 2\varepsilon/\sqrt{12}$. From (5), knowing that your estimate is the maximum of N such estimates, means that it is biased upwards by the amount $\sqrt{12}(N-1)\sigma/2(N+1)$. For $N = 2$–8, the resulting bias is quite similar to what would be found if the x_i were normally distributed around x_0.

[13] Kagel, Harstad, and Levin experimentally investigate first-price sealed bid auctions with positively affili-

where the expression Y is the same as in (2) above.

Comparing equation (5) with (6) shows that strategic discounting produces a winner's curse whenever $N > 3$. Further, strategic considerations (6), item valuation considerations (5), and the RNNE bid function (2) all call for greater discounting of bids relative to signal values with increases in ε.[14] Since evidence from private value auctions shows bidders to be sensitive to the strategic considerations inherent in these auctions (James Cox, et al.; our 1985 paper), we expect bidders to be sensitive to variations in ε.

However, with increasing numbers of bidders, equations (2), (5), and (6) give conflicting directions. Differentiation of (6) shows that strategic considerations require higher bids in the presence of more rivals as signal values are more congested. Item valuation considerations as expressed in (5) require less aggressive bidding as the adverse-selection problem becomes more severe. The net effect of these two forces, expressed in the RNNE bid function, is for bids to remain constant or decrease in the presence of more rivals.[15] This conflict between item valuation and strategic considerations suggests that structural variations in numbers of bidders is critical to determining whether experienced bidders learn to avoid the winner's curse in small groups out of a trial and error survival process that is situationally specific, as opposed to "understanding" the adverse-selection problem as it applies to new situations. If the survival process results in generalized learning (one interpretation of optimality via survival arguments in economics), behavior in both large and small groups should show comparable deviations from the RNNE reference point.[16] However, in the absence of generalized learning, behavior is likely to be markedly different, to the extent that the winner's curse, having been largely or entirely eliminated in small groups, will reemerge with increases in group size.

B. *Effects of Public Information*

1. *The Nash Equilibrium.* Extensions of the common value auction model show that public information reducing item valuation uncertainty will increase average seller's revenues (reduce bidder's profits) under the RNNE (Milgrom and Weber). This holds even though public information signals of the sort employed here, x_p, will on average lie below the maximum private information signal, x_1. The economic forces at work here are roughly as follows: On average,

$$E[x_0|X_i = x_1] = E[x_0|X_i = x_1, X_p]$$

for the bidder actually holding the highest private information signal, x_1. (All symmetric, noncooperative Nash equilibria involve agents bidding as if $x_i = x_1$, since their bid only "counts" when this presumption is satisfied.) However, due to the affiliation of the signal values, for bidders whose private information signals $x_i < x_1$, the public information signal will, *ex post*, raise the average expected value of the item. This will induce an upward revision of these bids, which in

---

ated private values, where private values were generated using exactly the same procedures generating the x_i values here. Equation (6) corresponds to the bid function developed there. Note, these private value experiments show bids commonly in excess of (6) with $N = 6$.

[14] In addition, an increase in ε increases the variance associated with the naive expectation (4). Thus to the extent that bidders are generally risk averse and bid more cautiously in the face of increased risk, but are poor Bayesians so that they continue to employ the naive expectation (4), they will respond correctly (at least directionally) to changing ε.

[15] Differentiation of (2) with respect to N shows the Y term to require lower bids in the presence of more rivals.

[16] An earlier reader of this paper argued that optimality by survival arguments implied "correct" responses under all possible states of the world, as survivors had experienced all relevant states and had learned to adapt to them. While acknowledging the validity of this interpretation, it seems to rob the survival argument of empirical content as: 1) the survival process would never be complete as real economies are repeatedly subject to changing conditions and changes in the set of agents, and 2) there would be no role left for economic theory in terms of understanding behavioral processes or in predicting responses to novel economic conditions.

turn puts pressure on the bidder with the highest private information signal, x_1, to bid more out of strategic considerations.

As an experimental device, the use of the low private information signal, x_L, as the public information signal has several advantages. First, the RNNE bid function with public information is readily solved analytically with x_L. Second, x_L provides a substantial amount of information concerning the location of x_0, and the signal values rivals are likely to have. For signals in the interval (1),

$$(7)\quad E\left[x_0|X_i=x_1, X_p=x_L\right]=(x_L+x_i)/2$$

provides a sufficient statistic for the value of x_0 given the set of private information signals distributed, under the presumption that $x_i=x_1$. From (7), it is clear that announcing x_L should reduce the average spread in beliefs about the underlying value of the item between any two bidders by one-half. Under private information conditions, a similar reduction in beliefs would require halving ε, as this halves the average spread between any two private information signals (as well as the spread in expected values under (5)). Thus, announcing x_L induces strong competitive pressures on the high bidder and translates into relatively large increases in seller's revenues, or reduced bidder's profits (see equation (9) below), while still maintaining an interesting auction.

The RNNE bid function with, public information, x_L, and private information signals in the interval (1) is[17]

$$(8)\quad b(x_i,x_L)=x_L+\left[\frac{N-2}{N-1}\right]\times\left[\frac{x_i+x_L}{2}-x_L\right]=\frac{N}{2(N-1)}x_L+\frac{(N-2)}{2(N-1)}x_i.$$

Expected profits are substantially diluted, being

$$(9)\quad E[\Pi|W,X_L]=\varepsilon/(N+1).$$

This is a little more than one-half of the expected profits under private information conditions (3).

With risk aversion we cannot unambiguously determine whether public information will increase or decrease average seller's revenues (Milgrom and Weber). The impact depends upon the particular form of the utility function assumed, the degree of risk aversion displayed, and the extent to which public information dilutes private information differences. Nevertheless, we would anticipate that under most plausible scenarios, the relatively large dilution of private information differentials inherent in releasing x_L, would cause seller's revenues to increase, or at least not to decrease.

2. *Judgmental Failures: The Winner's Curse.* The judgmental error underlying the winner's curse consists of the high bidder's systematic overestimation of the item's value. To the extent that the magnitude of these judgmental errors decreases as the uncertainty concerning the value of the item decreases, public information will result in a downward revision in the most optimistic bidder's valuation of the item. This introduces a potentially powerful offset to any strategic forces tending to raise bids.[18] This effect is well illustrated through extending the notion of strategic discounting to auctions with public information, and comparing the resultant discount function with (6).

---

[17] For x_i in the interval (1), the RNNE equilibrium bid function is obtained under the initial condition that $b(x_L)=x_L$. Note that ε is not explicitly represented in (8). However, the average difference $(x_i+x_L)/2-x_L$ depends directly on ε.

[18] Note that unbiased random errors, even if they do not result in high bids in excess of $E[x_0|X_i=x_1]$, will *not* cancel out here in terms of their effects on seller's revenues. This results from the auction selection mechanism whereby market outcomes overrepresent bids with upward biases, resulting in average bids in excess of the RNNE. Consequently, if public information reduces the magnitude of these item valuation errors, public information will still result in a downward revision of the market price, which offsets the strategic forces promoting increased revenues. High bids in excess of $E[x_0|X_i=x_1]$ can constitute an extreme form of these errors and/or a systematic tendency, on at least some bidder's part, to overestimate the item's value.

Under strategic discounting, we continue to assume that bidders employ naive expectations to determine the value of the item, to the point that they ignore the positional information inherent in announcing x_L, and act as if they are in a private value auction. For x_i and x_L in the interval (1), a naive expectation of x_0 is the same as (7):

$$E[x_0|x_i, x_L] = (x_i + x_L)/2. \tag{10}$$

Consequently, under risk neutrality, the strategic discount function and the RNNE bid function (8) coincide here. Note that this follows directly from the fact that the naive (10) and sophisticated expectations (7) coincide. In markets with private information, the two bid functions differ as the expectations differ.

Comparing (6) with (8), $b^s(x_i) > b(x_i, x_L)$ on average for all $N \geq 3$. In auctions where bidders employ naive expectations, but strategic discounting, announcing x_L will result in *reductions* in average seller's revenues (increases in average bidder's profits) with 3 or more bidders. Finally, given that previous experimental studies indicate sensitivity to the strategic implications inherent in auction markets, and the coincidence of the strategic bid function with the RNNE bid function, the RNNE model should provide a fairly good predictor of market performance with x_L announced, irrespective of its predictive adequacy in comparable markets with private information only.

C. *Summary of Research Questions of Primary Interest*

We conclude this section by summarizing the research questions of primary interest in the form of hypotheses to be tested.

HYPOTHESIS 1: *Under private information conditions market outcomes for experienced bidders are observationally indistinguishable from the RNNE as* (i) *the high signal holder usually wins the auction, and* (ii) *prices do not deviate substantially or systematically from the RNNE prediction.*

HYPOTHESIS 2: *Announcing x_L, the lowest private information signal, raises average seller's revenues by the average amount predicted under the RNNE model.*

HYPOTHESIS 3: *Under private information conditions, experienced bidders avoid the winner's curse as average profits are closer to the RNNE level than the zero/negative profits predicted under the winner's curse.*

HYPOTHESIS 4: *Public information raises average seller's revenue.*

HYPOTHESIS 5: *Hypotheses* 3 *and* 4 *apply uniformly to experiments with small and large numbers of bidders.*

HYPOTHESIS 6: *Bidders are sensitive to the strategic implications of the auctions so that when the strategic discounting model and the RNNE model coincide, the RNNE model provides a reasonable characterization of the data.*

Hypotheses 1 and 2 involve strong predictions, which if satisfied would imply satisfaction of all the other hypotheses as well. Hypotheses 3–5 involve weaker predictions which may be satisfied even though Hypotheses 1 and 2, strictly interpreted, fail. Nevertheless confirmation of these weaker predictions would indicate that the RNNE model provided a reasonable "ballpark" characterization of behavior in general: that for experienced bidders, at least, the repeated nature of market decision processes in conjunction with survival pressures and profit opportunities, eliminate the judgmental failures underlying the winner's curse. Further, conditional on the confirmation of Hypotheses 3–5, one of the key policy implications of the theory, the revenue-enhancing effects of public information, would be reasonably accurate as well.

Finally, Hypothesis 6 captures the notion that bidders in private value auction experiments have been shown to be sensitive to the strategic implications inherent in these auctions. Hence, we would expect the RNNE model to perform well here when its predictions coincide with strategic discounting. If Hypothesis 6 is satisfied, but Hypotheses 1–5 are not, we have indirect evidence that it is the judgmental errors underlying the

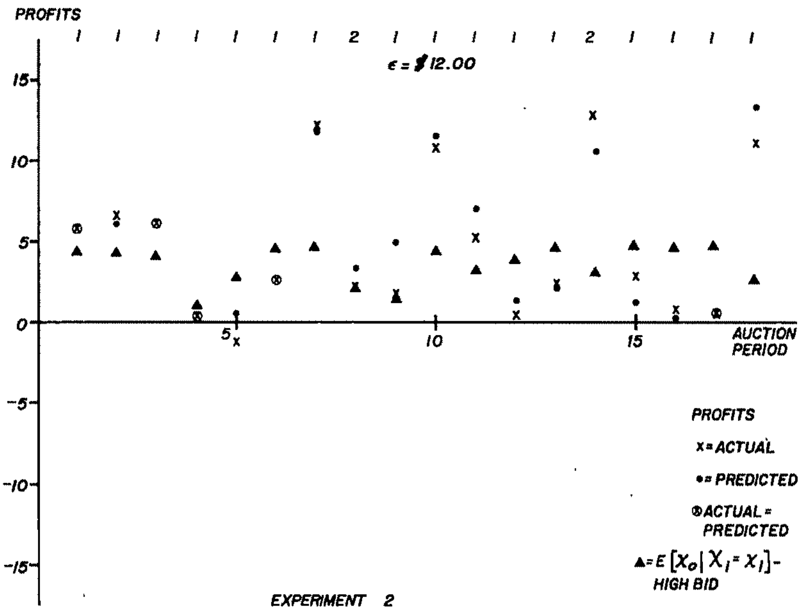


FIGURE 1

winner's curse that are responsible for the breakdown in the model's performance.

III. Experimental Results

A. *Bidding Patterns with Private Information*

Figures 1–5 provide representative data for market outcomes over time. The rank of the high bidder's signal value is shown at the top of each figure: 1 for the highest signal value, 2 for the second highest, etc. Closed triangles show the difference between $E[x_0|X_i = x_1]$ for the high bidder and the high bid. A negative (positive) value here indicates that the high bid exceeded (fell below) $E[x_0|X_i = x_1]$, implying negative (positive) expected profits should the high signal holder win the auction. Cross marks show actual profits earned, with closed circles showing profits predicted under the RNNE: cross marks below (above) closed circles indicate that the actual bid exceeded (was less than) the RNNE prediction and by how much. Eyeballing the data, there appears to be little systematic variation in bids over time, within a given auction series, independent of variations in N and ε.

Table 2 provides summary statistics of auction outcomes.[19] Columns 4–6 show, respectively, actual profits earned, profits predicted under the RNNE, and profits earned as a percentage of the RNNE prediction. For comparative purposes, column 3 shows profits predicted under the strategic discounting formulation (profits predicted using the bid function 6, assuming that the highest signal holder always won the auction). The last two columns report the average frequency with which the high signal holder won the item, and the frequency with which the high bid exceeded $E[x_0|X_i = x_1]$.

---

[19] Our computations include all auctions under dual market procedures, regardless of whether bidders actually made profits (or losses) as a consequence of our coin flip procedure. Since the coin flip was made after bids from both markets had been accepted, its outcome should not affect decisions. Experiments involving experienced, as opposed to super-experienced, subjects had one or more dry runs with no money at stake. These auction periods are *not* included in the analysis.

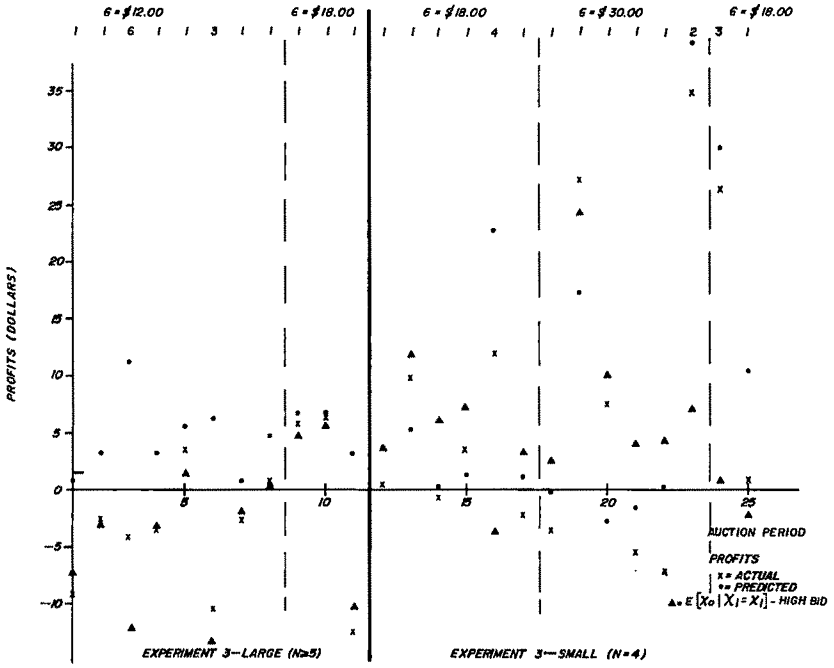


FIGURE 2

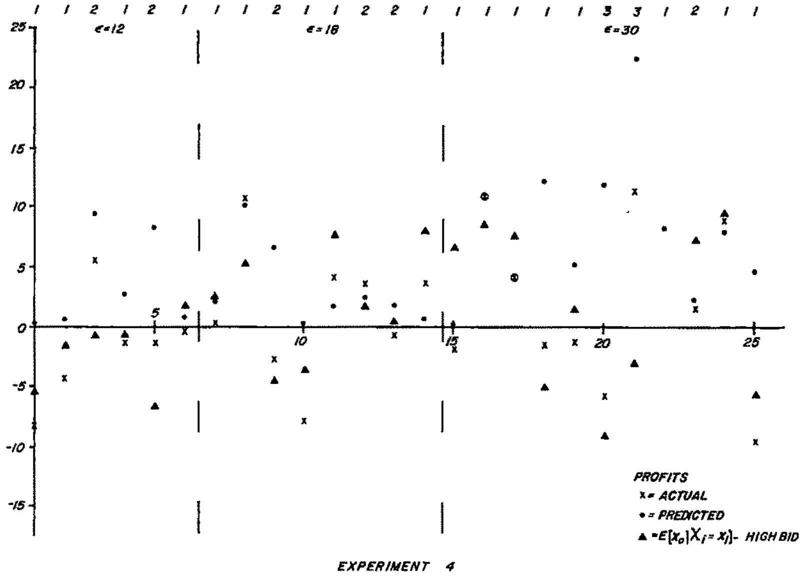


FIGURE 3

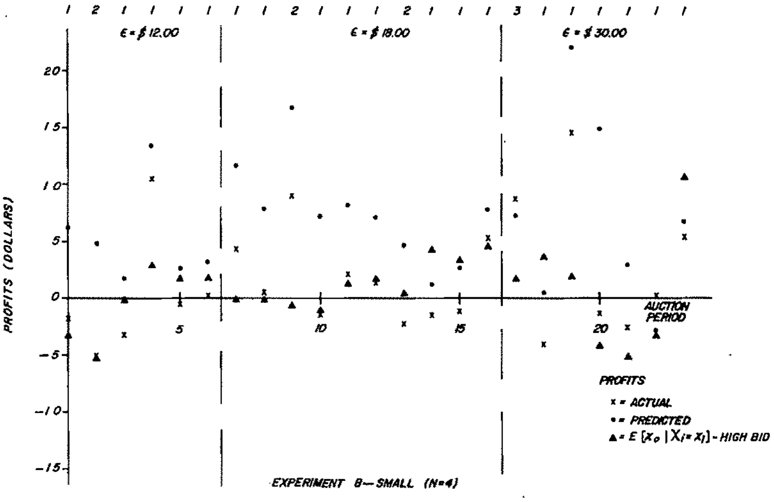


FIGURE 4

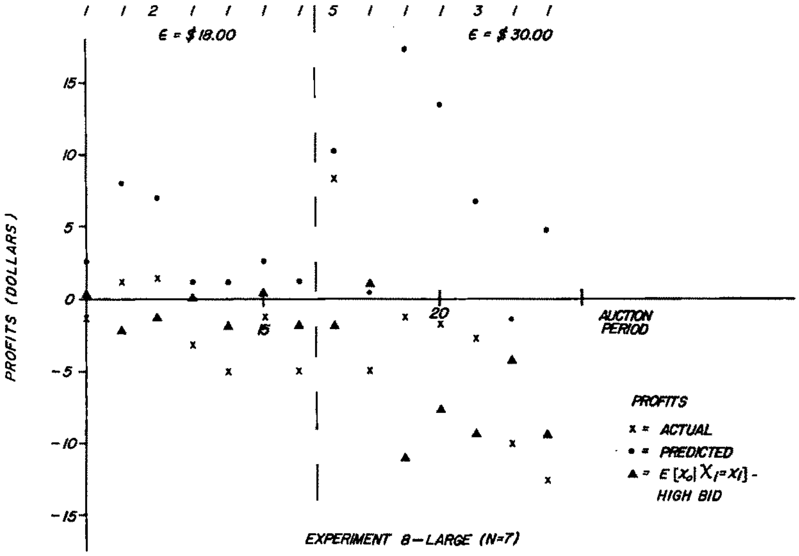


FIGURE 5

The auction series were ordered by the number of active bidders, beginning with the small group experiments, as we kept drawing the same conclusion: there were substantial differences in the ability of the RNNE bidding model to provide a ballpark characterization of the data in auctions involving small numbers (3–4) of bidders compared to those with large numbers (6–7) of bidders. In auction series that began with small numbers of bidders, one can observe bankruptcies (series 6) and some bidding in excess of $E[x_0|X_i = x_1]$. There was at least one large group series where average profits were positive and bidding was generally below $E[x_0|X_i = x_1]$ (series 7 large). However, the

TABLE 2—PROFITS AND BIDDING BY EXPERIMENT AND NUMBER OF ACTIVE BIDDERS: PRIVATE INFORMATION CONDITIONS (Profits measured in dollars)

Auction Series (No. of Periods)	No. of Active Bidders	Average Profits with Strategic Discounting (standard error of mean)	Average Actual Profits (t-statistic)[a]	Average Profits Under RNNE (standard error of mean)	Profits as a Percentage of RNNE Prediction	Percentage of Auctions Won by High Signal Holder	Percentage of High Bids $b_1 > E[x_0 \mid X_i = x_1]$
6 (31)	3–4	3.25 (1.51)	3.73 (2.70)[b]	9.51 (1.70)	39.2	67.7	22.6
2 (18)	4	−.75 (1.07)	4.61 (4.35)[c]	4.99 (1.03)	92.6	88.9	0.0
3 small (14)	4	−3.82 (2.40)	7.53 (2.07)	6.51 (2.65)	115.7	78.6	14.3
7 small (19)	4	−.12 (1.56)	5.83 (3.35)[c]	8.56 (2.07)	68.1	63.2	10.5
8 small (23)	4	−2.24 (1.05)	1.70 (1.56)	6.38 (1.21)	26.6	84.6	39.1
1 (18)	5	−1.90 (.85)	2.89 (3.14)[c]	5.19 (.86)	55.7	72.2	27.8
3 large (11)	5–7	−5.19 (.55)	−2.92 (−1.49)	3.65 (.62)	−80.5	81.8	63.6
7 large (18)	6	−10.11 (.96)	1.89 (1.67)	4.70 (1.03)	40.2	72.2	22.2
4 (25)	6–7	−10.03 (1.05)	−.23 (−.15)	4.78 (.92)	−4.8	69.2	46.2
5 (26)	7	−8.07 (1.04)	−.41 (−.44)	5.25 (1.03)	−7.8	42.3	65.4
8 large (14)	7	−11.04 (1.35)	−2.74 (−2.04)	5.03 (1.40)	−54.8	78.6	71.4

[a] Tests null hypothesis that mean is different from 0.0.
[b] Significant at 5 percent level, 2-tailed t-test.
[c] Significant at 1 percent level, 2-tailed t-test.

general pattern was one of positive average profits in small groups which, while well below the RNNE criteria, were clearly closer to the RNNE prediction than the zero/negative profit levels of the winner's curse: profits averaged across experiments with 3–4 bidders were $4.68 per auction period, about 65.1 percent of the RNNE models prediction of $7.19 per auction period.[20] In contrast, auctions involving 6–7 bidders had average actual profits of −$.88 per auction period. While this is substantially better than profits predicted under strategic discounting, −$8.89, indicating considerable adjustment to the adverse-selection problem, these adjustments were far from complete, as profit levels were negative, closer to the winner's curse prediction than the RNNE prediction of $4.68 per auction period. Further, comparing large and small group auctions, actual profits decreased substantially more than profit opportunities as measured by the RNNE criteria. The latter profit criteria dropped by $2.51 per auction period, while actual profits fell by $5.56. Thus, in going from small to large groups, profit perfor-

---

[20] Averages reported here and elsewhere in the text are simple, unweighted averages across experiments, unless noted otherwise. All RNNE profit calculations are exact, based on the bid function in (2) and fn. 8.

TABLE 3—PROFITS AND BIDDING UNDER VARYING LEVELS OF ε
(All profits in dollars)

No. of Bidders	ε	Average Profits with Strategic Discounting (standard error of mean)	Average Actual Profits (t-statistics)[a]	Average Profits Under RNNE[b] (standard error of mean)	Profits as a Percentage of the RNNE Prediction
3–4	12	−1.24	2.60	4.52	57.5
		(.69)	(1.74)	(1.44)	
	18	−.24	3.98	7.20	55.4
		(1.03)	(3.71)[d]	(1.05)	
	24/30	.60	6.75	11.22	60.2
		(1.94)	(3.33)[d]	(2.06)	
6–7	12	−3.68	−1.86	3.46	−53.8
		(.54)	(−2.21)[c]	(.56)	
	18	−8.51	−.95	3.19	−29.8
		(.52)	(−1.00)	(.51)	
	24/30	−12.31	.60	7.12	8.4
		(.89)	(.51)	(.94)	

[a] Tests null hypothesis that mean is different from 0.0.
[b] Based on sample of signal values drawn.
[c] Significant at 5 percent level, 2-tailed t-test.
[d] Significant at 1 percent level, 2-tailed t-test.

mance deteriorated above and beyond that predicted under the RNNE.

In both small and large groups, the bidder with the high private information signal generally won. Averaging across auction series, the percentages were 76.6 and 68.8 percent for small and large groups, respectively. Thus, the adverse-selection mechanism hypothesized to underly the winner's curse was present for both group sizes. The high bid, b_1, was below $E[x_0|X_i = x_1]$ in 82.7 percent of all auctions involving small number of bidders, but was below this value in only 46.2 percent of those auctions involving large numbers. The judgmental errors underlying the winner's curse were largely absent in small groups, but were quite prevalent in larger groups.[21]

Table 3 shows the relationship between profits and ε where the averages are computed over auction periods. We continue to distinguish between auctions involving small (3–4) and large (6–7) numbers of bidders, and pool the data for $\varepsilon = 24$ and $\varepsilon = 30$ as there were relatively few observations at $\varepsilon = 24$. In auctions involving small numbers of bidders, actual profits increased with increases in ε and constituted a relatively stable fraction of the profits predicted under the RNNE. With large numbers of bidders (6–7), average losses decreased with increases in ε, with positive average profits earned at $\varepsilon = 24$–30. Reduced losses with ε increasing implies that bids were reduced proportionately more relative to signal values. A stable proportionate discount in terms of ε would have resulted in larger absolute dollar losses as ε increased.[22] Instead, subjects took advantage of the increased profit opportunities inherent in the increased value of ε.

---

[21] The adverse-selection discount identifying the winner's curse fails to account for the high signal holder not always winning the item. Bidders should have been responsive to this and used a smaller adverse-selection discount based on actual frequencies with which different ranked signal holders won. Since over 90 percent of all auctions were won by the first- or second-highest signal holder, $E[x_0|X_i = x_2]$ serves as a reasonable *upper* bound on the relevant discount. Using this measure, less than 1 percent of all small group auctions had a winner's curse, while 28.7 percent of all large group auctions did. Our conclusion regarding the differential frequency of the winner's curse in large and small groups is unaffected, although the overall frequency is reduced substantially.

[22] Recall that $E[x_0|X_i = x_1]$ is decreasing in ε.

We used multiple regression analysis to summarize and quantify the influences on bidding in the private value auctions. Using an error components model, and restricting our analysis to signal values in the interval (1), Table 4 shows the results of two alternative bid function specifications.[23] The first regression involves a generalized version of the bid function (2), allowing for a nonzero intercept and including numbers of bidders as a right-hand side variable. Under this specification, both the intercept coefficient and the coefficient associated with the variable Y in (2) were not significantly different from zero. The second regression drops these two terms. Under both specifications, the x_i critically influences bids with an estimated coefficient value close to 1.0, as implied under the RNNE bid function and under strategic bidding. Further, bidders were clearly sensitive to changes in ε, so that we can soundly reject any naive bidding models which postulate a constant, fixed discount relative to x_i. On the other hand, the coefficient on ε is significantly below unity, and this contradicts the RNNE bid function. These results parallel those from private value auctions, where increases in the distribution of underlying values resulted in lower bids, but these bids were not reduced as much as predicted under the risk neutrality hypothesis (Kagel, Harstad, and Levin). Finally, the aggressive forces associated with increased numbers of bidders win out over the item valuation influences, as we find a statistically significant, positive coefficient associated with N under both specifications. The increased aggressiveness of individual bids with increases in N is in direct contradiction to the predictions of the RNNE bidding model. It adds to the dilution of bidders profits associated with increased N predicted under the RNNE, with the net result a persistent winner's curse with $N = 6$ or 7.

TABLE 4—ERROR COMPONENTS ESTIMATES OF BID FUNCTION IN PRIVATE INFORMATION MARKETS[a]

$$\begin{array}{lllllr}
b(x_i) = & -\ 4.30 + & 1.00x_i - & .73\varepsilon & & \\
 & (3.36) & (.002) & (.03) & & \\
 & & +\ \ .70N - & .02Y & & R^2 = .99 \\
 & & (.16) & (.14) & & \sigma_\eta = 4.94 \\
b(x_i) = & 1.00x_i - & .74\varepsilon + & .65N & & R^2 = .99 \\
 & (.002) & (.03) & (.15) & & \sigma_\eta = 4.94
\end{array}$$

[a] Standard errors are shown in parentheses.

Learning patterns over time are not explored in detail here. The interested reader should consult Kagel et al. (1986) where individual learning patterns of inexperienced bidders are analyzed. The data reported there show considerable learning on survivors part, as most start out bidding in excess of $E[x_0|X_i = x_1]$. Comparing the results here with end-of-experiment performance for these inexperienced bidders shows continued learning as well: averaging over the last 5 periods of 5 auction series of 3–4 inexperienced bidders shows 28 percent of all high bids in excess of $E[x_0|X_i = x_1]$ (compared to 17.3 percent here) with profits averaging 27.5 percent of the RNNE prediction (compared to 65.1 percent here).[24] It is important to note, however, that the learning resulting in the strong performance of the RNNE model in small groups appears to be situationally

---

[23] An error components specification was employed, with error term

$$\eta_{it} = u_i + v_{it} \qquad i = 1,\dots,N; \quad t = 1,\dots,T;$$

where u_i is a subject-specific error term, assumed constant subject across auction series, and v_{it} is an auction period error term. Standard assumptions were employed: $u_i \sim (0, \sigma_u^2)$ and $v_{it} \sim (0, \sigma_v^2)$ where u_{it} and v_{it} are independent among each other and among themselves. Badi Baltagi's (1986) weighted least squares computational procedure was used to invert the variance-covariance matrix. A fixed-effects error specification generated similar coefficient estimates and standard errors. Permitting the u_i to vary with subject participation in different auction series yields similar estimates except for the variable N, which increases in value with no loss in statistical significance.

[24] There are no large group end-of-experiment data from the inexperienced subject experiments to compare with the results reported here, as bankruptcies precluded keeping groups of 6–7 bidders intact for very long.

TABLE 5—EFFECTS OF PUBLIC INFORMATION ON SELLER'S REVENUES
(All revenue figures in dollars)

		Average Change in Revenues		
		Actual (*t*-statistic)[a]	Predicted (standard error of mean)	
Acution Series (No. of Periods)	No. of Active Bidders		RNNE Model	Strategic Discounting
6	3–4	4.38	7.62	1.38
(23)		(2.71)[b]	(1.54)	(1.41)
3 small	4	−2.71	3.83	−6.51
(14)		(−.99)	(2.05)	(1.58)
7 small	4	6.58	7.46	−3.46
(11)		(2.79)[b]	(1.74)	(1.52)
7 large	6	.22	1.64	−13.16
(18)		(.15)	(.87)	(1.26)
4	6–7	−3.20	1.25	−15.75
(18)		(−1.36)	(.82)	(1.07)
5	7	−2.40	2.43	−12.34
(20		(−1.83)	(.86)	(1.30)

[a] Tests null hypothesis that mean is 0.0.
[b] Significant at 5 percent level, 2-tailed *t*-test.

specific, and does not provide the "understanding" to respond appropriately to increased numbers of rivals (see Table 4), or to avoid the winner's curse in large group auctions.

B. *Effects of Public Information on Seller's Revenues*

Table 5 shows the actual effects of announcing x_L on average revenues and the predicted effects under both the RNNE and the strategic discounting models. As in Table 2, we have ordered the results by number of active bidders.

Averaging across the 3 auction series with small numbers of bidders, x_L raises revenues an average of \$2.75 per auction period. Pooling auction periods, a *t*-test shows this to be statistically significant ($t = 2.41, p < .02$, 1-tailed test), although well below the RNNE prediction of a \$6.30 increase in revenues. Public information reduced revenues in series 3, the auction series which came closest to the RNNE model's predictions under private information conditions. Nevertheless, this seems like an anomaly rather than the norm, since under all other conditions where the winner's curse was weak or nonexistent, as it was in series 6 and series 7 small and large, revealing the low information signal, x_L, raised seller's revenues (also see the discussion that follows).

Averaging across auctions with large numbers of bidders, public information *reduced* seller's revenues \$1.79 per auction period. A *t*-test based on pooled observations indicates that this decrease is statistically significant ($t = -1.48, p < .08$, 1-tailed test). While the reduced revenues are well below the predictions of the strategic discounting model, they are opposite in sign and magnitude to the RNNE model which calls for a \$1.78 *increase* in revenue.

A within-auction series analysis reinforces the importance of the winner's curse in determining whether public information raises or lowers revenues. The first two columns of Table 6 show the effects of announcing x_L conditional on the presence or absence of a winner's curse in large group auctions, for those auction periods where the RNNE model predicts an increase. A *chi*-square test shows the winner's curse to significantly affect the validity of the RNNE model's prediction ($\chi^2 = 4.25, p < .05$). The third

TABLE 6—EFFECTS OF THE WINNER'S CURSE ON REVENUE RAISING EFFECTS OF PUBLIC INFORMATION[a]

	Number of Periods in Auction		
	Large Numbers (6–7)		Small Numbers[b] (3–4)
Change in Seller's Revenues	Winner's Curse	No Winner's Curse	No Winner's Curse
Increase	6	12	29
Decrease	11	5	10

[a] Auction market periods where RNNE predicted an increase in seller's revenues. Winner's curse defined in terms of high bid in private information market in excess of $E[x_0|X_i = x_1]$.

[b] Winner's curse present in 3 auction periods where RNNE predicts increase in seller's revenues. Hence omitted.

column of the table shows the effects on revenue in the absence of a winner's curse in the small group auctions. The pattern here is much the same as found in auctions with 6–7 bidders in the absence of a winner's curse ($\chi^2 = .086$).[25]

In the two auction series where public information consisted of posting an additional, randomly drawn, private information signal, public information raised revenues \$.27 per auction (an average increase of −\$.80 in experiment 1 and \$1.33 in experiment 2). Note that neither of these series exhibited significant traces of the winner's curse under private information conditions. In contrast, in an earlier series of experiments involving inexperienced subjects, revealing a random public information signal consistently *reduced* revenues, in this case by an average of \$2.95 per auction (Kagel et al., 1984). In 2 of these 4 auction series, average profits were negative under private information conditions, while in 3 of the 4, the high bid exceeded $E[x_0|X_i = x_1]$ in 40 percent or more of the auctions. These results serve to reinforce the hypothesis that an absence of a winner's curse is a necessary condition for public information to raise average revenues.

Table 7 reports market outcomes relative to the RNNE model's predictions with x_L announced. Average actual profits were positive in all experiments. While there is considerable variation in profits relative to the RNNE model's predictions across auction series (especially series 3 and 6), on average profits were only slightly less than predicted (\$3.20 actual vs. \$3.41 predicted). Further, unlike private information conditions, there were no systematic differences in realized profits relative to predictions as the number of active bidders varied. These results, at the market level at least, are consistent with our earlier suggestion that the RNNE model would provide a more accurate characterization of performance with x_L announced, compared with private information conditions.

More detailed analysis of the data, however, shows that more is at work here than simple strategic discounting with all bidders employing identical bid functions. With x_L announced, there is almost a complete breakdown of the prediction that the bidder with the highest private information signal wins the auction. In these experiments the high private information signal holder won only 29.5 percent of all auctions. This is only modestly above what one would expect if chance factors alone determined whether the high private signal holder won, an expected frequency of 21.8 percent.

---

[25] Results similar to these have been found in a companion series of 6 second-price common value auction experiments.

TABLE 7—PROFITS AND BIDDING WITH PUBLIC INFORMATION (x_L)
(All profits in dollars)

Auction Series (No. of Periods)	No. of Bidders	Average Actual Profits (t-statistic)[a]	Average Profits Under RNNE (standard error of mean)	Profits as a Percentage of RNNE Prediction	Percent of Auctions Won by High Signal Holder
6 (23)	3–4	.15 (.08)	2.96 (1.60)	5.1	30.4
3 small (14)	4	10.24 (3.33)[c]	2.68 (2.03)	382.0	21.4
7 small (11)	4	2.07 (.86)	4.54 (2.00)	45.6	18.2
7 large (18)	6	1.67 (1.56)	3.06 (.62)	54.6	27.8
4 (18)	6–7	3.43 (2.24)[b]	4.14 (1.04)	82.9	44.4
5 (20)	7	1.64 (1.62)	3.06 (.86)	53.6	35.0

[a] Tests null hypothesis that mean is 0.0.
[b] Significant at 5 percent level, 2-tailed t-test.
[c] Significant at 1 percent level, 2-tailed t-test.

Detailed examination of the data shows a handful of bidders (20.0 percent) winning a disproportionately large number (57.7 percent) of the auctions with x_L announced. This handful of bidders did quite well as a group, earning average profits of \$2.79 per auction period won, as compared to \$2.89 for all other bidders. A distinguishing characteristic of these bidders is that they were relatively more aggressive than their rivals under private information conditions (ranking bids as a fraction of signal values consistently placed them in the top half of all bidders). The increased aggressiveness of this handful of bidders in the public information markets (they won only 15.1 percent of all private information auctions in which they did not have the highest private information signal, compared to 54.7 percent of the corresponding public information auctions) is directly attributable to the sharp reduction in beliefs about the underlying value of item inherent in announcing x_L. This reduction in the effective dispersion of information concerning x_0 permitted differences in risk attitudes and information processing capacities to play an increased role in the outcomes of auctions with public information.[26]

Table 8 reports the results of statistical estimates of individual bid functions with x_L announced. Recall from (8) that the parameters associated with the variables x_L and x_i are a function of the number of bidders present. The specifications in Table 8 employ different slope coefficients for these variables in small group ($N = 3$–4) and large group ($N = 6$–7) cases. With $N = 3$–4, the estimated slope coefficients for public and private information, are close to the theoretical bid function prediction (with $N = 4$, these are .67 and .33 for public and private information, respectively). With $N = 6$–7, more weight is attached to public rather than

---

[26] However, these bidders were unable to overcome the inherent disadvantage of holding only public information, as their average profits did not deviate significantly from zero in cases where they held the low private information signal.

TABLE 8—ERROR COMPONENTS ESTIMATES OF BID FUNCTION IN MARKETS WITH x_L ANNOUNCED

$$b(x_i, x_L) = -\underset{(6.2)}{.95} + \underset{(.05)}{.24x_i} - \underset{(.06)}{.07x_i^*} + \underset{(.05)}{.72x_L} + \underset{(.06)}{.12x_L^*} + \underset{(.07)}{.14\varepsilon} + \underset{(.70)}{.08N} \qquad \begin{array}{r} R^2 = .98 \\ \sigma_\eta = 7.39 \end{array}$$

$$b(x_i, x_L) = \underset{(.05)}{.24x_i} - \underset{(.05)}{.07x_i^*} + \underset{(.05)}{.72x_L} + \underset{(.06)}{.12x_L^*} + \underset{(.06)}{.14\varepsilon} \qquad \begin{array}{r} R^2 = .98 \\ \sigma_\eta = 7.37 \end{array}$$

Note: $x_i^* = x_i$ if $N = 6$ or 7, $x_i^* = 0$ otherwise; $x_L^* = x_L$ if $N = 6$ or 7, $x_L^* = 0$ otherwise.

[a] Standard errors are shown in parentheses.

private information, in contrast to the theoretical bid function prediction (with $N = 7$, predicted weights are .58 and .42, respectively). Further, ε has a modest positive, statistically significant, effect on bids, contrary to the predictions of both the RNNE and strategic discounting models, in which x_L and x_i capture all the information necessary (recall equation (8)).

C. *Summary of Experimental Outcomes of Primary Interest*

The data permit us to reach clear conclusions regarding the research hypotheses of primary interest specified in Section II, subsection 3. With respect to Hypotheses 1 and 2, we reject the strong form of the RNNE bidding model, irrespective of the number of rivals in the market or whether the market involves private or public information. Bidding consistently exceeded the RNNE prediction in private information markets, was highly variable relative to the RNNE reference point in markets with public information, and x_L failed to raise revenues by the predicted amount, even in markets without a winner's curse. The weak form of the RNNE model consistently outperformed the winner's curse and the strategic discounting model in markets with small numbers of rivals, consistent with Hypotheses 3 and 4. However this ballpark characterization of the data failed on both counts in markets with large numbers of rivals, leading us to reject Hypothesis 5. Finally, bidders were sensitive to the strategic implications of the auctions, responding correctly to variations in ε and coming close, on average, to the RNNE model's prediction with x_L announced. This confirms Hypothesis 6, which in turn suggests that the rejection of Hypotheses 1–5, particularly the rejection of Hypothesis 5, follows from the judgmental errors underlying the winner's curse.

IV. Towards Generalizability: But Is This How the Real World Operates?

A common criticism of experimental research in economics is that behavior in the laboratory is unlikely to be representative of field behavior. This criticism increases, as well it should, with the degree to which laboratory behavior deviates from accepted economic theory and common understanding of what constitutes "rational" economic behavior. Critics argue that auction market subjects, MBA students and senior undergraduates, are inherently less sophisticated, and clearly less experienced, than executives in the relevant industry, that experimental subjects do not have as much time to think and respond to events as industry executives, and that they lack the assistance of expert advisors that many industries have, to cite some of the prominent criticisms we have encountered. One can rebut these arguments on grounds that the experimental designs drastically simplify the decision-making structure, thereby obviating the need for expert advisors and reducing time requirements to make sensible decisions. Further, experimental subjects receive substantially more feedback, with shorter delays, regarding the outcomes of their decisions, so that the feedback loops that promote learning and adjustment over time are substantially stronger for experimental subjects compared to industry executives. (This is particularly true in the case of OCS lease sales where executives know that returns on investment will only be revealed years after bids have been accepted, and the responsible parties might well be in different positions within the company, or moved to a rival firm.)

Logical arguments can only go so far in debates of this sort. The question posed here is, what do the relevant data outside the laboratory look like compared to laboratory-generated data? Is the same model capable of organizing behavior in both settings? Do the field data obviously contradict the laboratory data?

The remainder of this section examines these issues in the context of the U.S. government's outer continental shelf lease sales. Note that our objective here is not to definitively test between competing explanations using field data. If we thought the field data had this kind of potential, there would be no need to resort to laboratory experiments in the first place (see Vernon Smith, 1982, and Charles Plott, 1982, for general discussions of the problems involved in using field data to test between models of market behavior and the advantages of laboratory experiments). Rather, our objective is to show that a reasonable analysis of the available data does not falsify the hypothesis that similar economic processes are at work in both settings. If this can be done, the burden of proof rests on those who would argue that the results don't generalize to demonstrate that their arguments are correct.

The concept of a winner's curse arose from petroleum geologists analysis of OCS bidding patterns and industry based calculation of rates of return from winning lease sales (Capen et al.; Lorenz and Dougherty). In a more recent analysis, Walter Mead, Asbjorn Moseidjord, and Philip Sorensen (1983) found after-tax rates of return on all OCS leases in the Gulf of Mexico issued from 1954 to 1969 to be less than average returns on equity for U.S. manufacturing corporations. Mead et al. view lease purchases as high risk investments and conclude that "...they (lessees) have historically received no risk premium and may have paid too much for the right to explore for and produce oil and gas on federal offshore lands" (1983,pp. 42–43).[27] In light of the effects of public information on bidder's profits reported here, a second element of Mead et al.'s (1983) calculations, namely rate of return *differentials* between drainage and wildcat leases, provides important corroborating evidence for the argument that lessees probably paid "too much" in these early OCS lease sales. The remainder of this section details this argument.[28]

A wildcat lease involves a tract for which there are no drilling data available that would indicate potential productivity. When a hydrocarbon reservoir has been located on a wildcat tract that is expected to extend into adjacent unleased acreage, the adjacent tract is defined by the U.S. Geological Service (USGS) as a drainage tract. Considerably more information is available regarding the economic potential of drainage than wildcat tracts. This information has both public and private components. An important public information component is that drainage tracts are unlikely to be dry, thereby significantly reducing the uncertainty (relative to wildcat tracts) that hydrocarbons will be found. However, developers of the wildcat tract (called neighbors) are likely to have superior private information relative to non-neighbors regarding the quantity of oil likely to be found, oil pressures, and other significant seismic information.[29]

---

[27] There is some argument as to whether investors require risk premiums for investing in oil and gas leases. A number of writers suggest that risk-averse bidders would require a premium relative to investing in alternative activities. Others, one of our referees included, argue that large oil companies with access to capital markets and having a diversified portfolio of leases would not be expected to earn risk premiums.

[28] Our analysis is not concerned with absolute rates of return, or absolute present discounted value calculations, for OCS sales compared to other industries. Rather, we are concerned exclusively with *differences* in rate of return between drainage and wildcat leases. Differential rate calculations within an industry, by the same research team, should be relatively more robust to the empirical problems encountered in obtaining such measure than comparisons across industries by different research groups.

If the information available on drainage leases were purely public, it should, according to Nash equilibrium bidding theory, raise average seller's revenues, hence reducing bidder's profits (recall Section II).[30] If the information were purely private, under Nash equilibrium bidding theory it would increase the rate of return for insiders (neighbors) relative to outsiders (nonneighbors), *and* reduce the average rate of return for nonneighbors below what would be earned in the absence of insider information (Wilson, 1975a, b; M. Weverbergh, 1979). If the added information on drainage leases contains both public and private information elements, rates of return for neighbors should be greater than for nonneighbors, but with nonneighbor returns definitely less than in the absence of the additional information (both the public and private information components push in this direction for nonneighbors).

What Mead et al. found were higher rates of return on drainage compared to wildcat leases for *both* neighbors (88.6 percent higher) *and* nonneighbors (56.2 percent higher). Further, nonneighbors won 43.2 percent of all drainage leases. While the higher rate of return for neighbors compared with nonneighbors can be explained by the presence of insider information (the explanation Mead et al. offer, 1983, 1984), the substantially higher rates of return for nonneighbors remains puzzling within the context of Nash equilibrium bidding theory. However, the higher rate of return for *both* neighbors and nonneighbors on drainage leases is perfectly consistent with our experimental findings, given the existence of a winner's curse in bidding on wildcat leases. According to this explanation, the additional information available from neighbor tracts served to correct for the overly optimistic estimate of lease value recorded in the average winning bids on wildcat tracts, thereby raising average profits for both neighbors and nonneighbors alike. In this respect, the OCS lease data parallel our experimental results with public information in the presence of a winner's curse.

What alternative explanations are available to explain *both* nonneighbors and neighbors rates of return being higher on drainage leases? Mead et al. suggest two alternatives. First, one might argue that the lower rate of return on wildcat leases reflects the option value of the private information consequent on discovering hydrocarbons. The higher rate of return of neighbors over nonneighbors on the drainage leases certainly suggests that neighbors had valuable proprietary information, the prospect of which would depress the value of the wildcat leases. However, returns to this proprietary information were far from certain to be realized, while the differential overall rate of return on leases in the Gulf, counting wildcats alone vs. counting wildcats plus drainage leases, was small, amounting to 7 percent of the wildcat rate of return (Mead et al., 1983). Thus the revealed value of the option is small and is unlikely to fully account for the depressed rate of return on wildcats relative to nonneighbors. Second, one can argue that the existence of insider information (and its common knowledge) scared off nonneighbors so that they did not bid, or bid very little, relative to lease value. Consequently, when nonneighbors won, since they bid quite low, they obtained higher rates of return as well. However, the frequency with which nonneighbors won drainage leases seems inconsistent with this argument.[31] To be sure, the average number of bids on drainage leases was less than on wildcats (2.88 vs. 3.33), but the data suggest only a modest 6.4 percent decline in the rate of return in going from leases with 2 bids to leases with 3–4 bids (Mead et al., 1983; 1984).[32]

---

[29] Only drainage leases have neighbors, namely those responsible for the development of the neighboring wildcat tract.

[30] All leases had the same royalty rate and were allocated on the basis of a first-price cash bonus bid. Drainage leases were spread throughout the Gulf so that each lease is likely to represent an independent pool of oil (Asbjorn Moseidjord, personal communication). The revenue-raising (profit-reducing) effects of public information in the RNNE model are expectations based on samples of independent observations (Milgrom and Weber). Hence, the drainage lease sample satisfies the assumptions of the model.

[31] Under Nash equilibrium bidding theory, it is not perfectly clear what ought to happen to the frequency

Note that we do not dispute Mead et al.'s (1984) argument that neighbors had proprietary information leading to higher rates of return on drainage leases than wildcats, or on drainage leases relative to nonneighbors. What we are claiming is that this proprietary information does not fully account for the substantially higher rates of return on drainage leases over wildcat leases for *both* neighbors and nonneighbors alike. Rather this element of the data is more readily explained by the public information component of the drainage lease designation, in conjunction with a winner's curse on wildcat leases. This explanation has the virtue of parsimony and consistency with the experimental results reported here.[33]

Our analysis of field data has been limited to 1954–69, prior to the publication of Capen et al.'s article alerting the industry to the presence of a winner's curse, and suggesting ways to avoid it. Many would argue that adjustments in bidding in the 1970's, partly in response to Capen et al.'s article and related publications, has eliminated the winner's curse in OCS lease sales, although opinions are not unanimous on the subject (see Lorenz and Dougherty, for example). We know of no rate-of-return studies on drainage vs. wildcat leases for the 1970's similar to Mead et al.'s for the 1960's that might help resolve the issue. We do know that at times there are significant discrepancies between cognitive understanding of the "right" thing to do and actual behavior. For mineral rights auctions, this involves firms recognizing and admitting that their geologists' estimates of value (and their economists' estimates of future price) have a significant error component, and that in the absence of insider information are unlikely to be better (on average) than their rivals. Such an admission is no small matter when paying substantial salaries to these professionals. As such we reserve judgment for the moment on the issue of a continuing winner's curse in OCS sales. Even assuming elimination of a winner's curse in more recent OCS lease sales does not affect our argument here, however: the available data outside the laboratory are consistent with data inside it

---

with which the informationally disadvantaged will win auctions. We suspect that this depends critically on the underlying distributions of item value and private information signals, and the nature of the insider information. However, all formal Nash equilibrium bidding models developed to date have the less informed earning lower profits than under symmetric information conditions (Wilson, 1975a,b; Weverbergh).

[32] Numbers of bidders in field environments is endogenous, depending in part on perceived lease value, rather than exogenous, as commonly treated in the auction market literature (and as one can arrange for in the laboratory). As such it is far from clear why, in theory, rates of return should vary systematically with numbers of bidders in field environments, unless again we postulate the existence of a winner's curse that is exaggerated with increased numbers of bidders.

[33] Drainage tracts have sharply reduced exploration costs as there are substantially fewer dry holes drilled per lease than on wildcats. Further, there are reduced production costs as a consequence of existing investments on neighbor leases and possibilities of joint production. In efforts to reconcile Mead et al.'s (1983) estimates of higher rates of return on drainage leases with their own estimates that prior drilling raised seller's revenues in the Gulf, Jeffrey Leitzinger and Joseph Stiglitz (1984) argue that developers capture at least some of the rent associated with reduced production costs on drainage leases. No explanation is offered for how this can plausibly account for the full differential rate of return between wildcat and drainage leases. Nor why, since these savings are public knowledge and contain a strong common value element, traditional motions of rent capture in competitive markets should fail. (Reduced production costs are in large measure available to both neighbors and nonneighbors as a consequence of the federal government's ability to force unitization, and the strong effects of these enforcement powers on voluntary unitization of tracts; see Gary Liebcap and Steven Wiggins, 1985.) An alternative explanation is that Mead et al.'s rate of return estimates are incorrect. However, there are equally strong, if not stronger, reasons to suppose that Leitzinger and Stiglitz's estimates of information externalities, which are based on the size of the bonus bid on drainage compared with wildcat leases, are highly exaggerated, as the public information component associated with the drainage lease designation is systematically biased towards raising the expected value of these leases.

in the absence of extensive efforts to alert bidders to the presence of a winner's curse.

V. Conclusions

Our experiments provide an empirical example of a market where individual judgment errors significantly alter market outcomes. Bidders in common value auctions, as in other auctions, are sensitive to the strategic opportunities inherent in the auction process. However, when strategic considerations and adverse-selection forces resulting from uncertainty about the value of the item conflict, behavior fails to conform, in important ways, with the requirements of Nash equilibrium bidding strategies.

Although we reject the general applicability of Nash equilibrium bidding models, market outcomes come closer to the risk-neutral Nash equilibrium model's predictions than to the winner's curse in auctions with small numbers (3–4) of bidders. In addition, there is considerable adjustment to the adverse-selection problem with large numbers (6–7) of bidders. However, these adjustments are far from complete, as profit levels are consistently negative, in conformity with the winner's curse. The existence of a winner's curse in large groups, in conjunction with the positive effect of the number of bidders on the size of individual bids, indicates that avoidance of the winner's curse in small groups is specific to the situation, and does not carry over to auctions with larger numbers of bidders. Bidders have learned to avoid the winner's curse in small groups out of a trial and error survival process, as opposed to "understanding" the adverse-selection problem as it applies to new situations.

Accounting for judgmental errors in these markets has some practical policy implications in terms of whether sellers choose to obtain and release information narrowing down the value of the auctioned item. In the absence of judgmental errors, this information clearly enhances seller's revenues, as Nash equilibrium bidding theory predicts. In the presence of judgmental errors, however, such information will almost surely reduce average revenues, a factor ignored to date in the literature.

Given sufficient experience and feedback regarding the outcomes of their decisions, we have no doubt that our experimental subjects, as well as most bidders in "real world" settings, would eventually learn to avoid the winner's curse in any particular set of circumstances. The winner's curse is a disequilibrium phenomenon that will correct itself given sufficient time and the right kind of information feedback.[34] Clearly the attenuation of the feedback loop between a decision and determining the outcomes of that decision, as is commonly the case in outer continental shelf lease sales and a number of other settings, serves to perpetuate the phenomena. Further, comparative evaluations of management performance in terms of money "left on the table" (the difference between the high bid and the second high bid) can do little to arrest the problem since the winner's curse is ubiquitous and applies fairly uniformly across individuals at early stages of the learning process (Kagel et al., 1986). Finally, to the extent that market participants feel that they have an inside edge, and better judgmental abilities than their rivals, the winner's curse is bound to be difficult to eliminate.

Apart from the important task of replicating our experiments, a number of interesting research questions remain to be explored. Since avoidence of the winner's curse involves a learning process, exactly what mechanisms, if any, insure market memory of past mistakes? To what extent do new entrants learn from experience compared to learning from observation or formal education? Do inexperienced bidders learn more quickly in markets dominated by experienced bidders? What are the dynamics of markets characterized by continual entry of new "suckers" who must learn from personal experience, and do we observe these dynamics in field environments?

---

[34] Auction series 7 large clearly indicates this as it involved super-experienced subjects all of whom had been in at least two previous large group series.

INSTRUCTIONS

This is an experiment in the economics of market decision making. The National Science Foundation has provided funds for conducting this research. The instructions are simple, and if you follow them carefully and make good decisions you may earn a CONSIDERABLE AMOUNT OF MONEY which will be PAID TO YOU IN CASH at the end of the experiment.

1. In this experiment we will create a market in which you will act as buyers of a fictitious commodity in a sequence of trading periods. A single unit of the commodity will be auctioned off in each trading period. There will be several trading periods.

2. Your task is to submit written bids for the commodity in competition with other buyers. The precise value of the commodity at the time you make your bids will be unknown to you. Instead, each of you will receive information as to the value of the item which you should find useful in determining your bid. The process of determining the value of the commodity and the information you will receive will be described in Sections 6 and 7 below.

3. The high bidder gets the item and makes a profit equal to the difference between the value of the commodity and the amount they bid. That is,

(VALUE OF ITEM) − (HIGHEST BID) = PROFITS

for the high bidder. If this difference is negative, it represents a loss.

If you do not make the high bid on the item, you will earn zero profits. In this case, you neither gain nor lose money from bidding on the item.

4. You will be given a starting capital credit balance of \$10.00. Any profit earned by you in the experiment will be added to this sum, and any losses incurred will be subtracted from this sum. The net balance of these transactions will be calculated and paid to you in CASH at the end of the experiment.

The starting capital credit balance, and whatever subsequent profits you earn, permit you to suffer losses in one auction to be recouped in part or in total in later auctions. However, should your net balance at any time during the experiment drop to zero (or less), you will no longer be permitted to participate. Instead we will give you your participation fee and you'll be free to leave the auction.

You *are* permitted to bid in excess of your capital credit balance in any given period.

5. During each trading period you will be bidding in a market in which *all* the other participants are also bidding. After all bids have been handed in they will be posted on the blackboard. We will circle the high bid and note the second high bid, and post the value of the item. We will also indicate whether a profit or loss was earned by the high bidder.

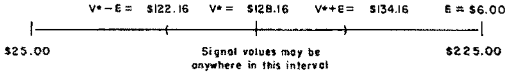
6. The value of the auctioned commodity (V^*) will be assigned randomly and will lie between \$25.00 and \$225.00 inclusively. For each auction, *any value* within this interval has an *equally likely chance* of being drawn. The value of the item can never be less than \$25.00 or more than \$225.00. The V^* values are determined randomly and independently from auction to auction. As such a high V^* in one period tells you nothing about the likely value in the next period—whether it will be high or low. It doesn't even preclude drawing the same V^* value in later periods.

7. Private Information Signals:

Although you do not know the precise value of the item in any particular trading period, you will receive information which will narrow down the range of possible values. This will consist of a private information signal which is selected randomly from an interval whose lower bound is V^* minus epsilon (ε), and whose upper bound is V^* plus epsilon. *Any value* within this interval has an *equally likely* chance of being drawn and being assigned to one of you as your private information signal. You will always know what the value of epsilon is.

For example, suppose that the value of the auctioned item is \$128.16 and that epsilon is \$6.00. Then each of you will receive a private information signal which will consist of a randomly drawn number that will be between \$122.16 ($V^* - \varepsilon = \$128.16 - \$6.00$) and \$134.16 ($V^* + \varepsilon = \$128.16 + \$6.00$). Any number in this interval has an equally likely chance of being drawn.

The line diagram below shows what's going on in this example.



The data below show the entire set of signals the computer generated for our sample bag. (Note we've ordered these signal values from lowest to highest.)

$V^* = \$128.16$; $\varepsilon = \$6.00$. Signal values: \$122.57
124.14
124.68
126.76
128.84
129.51
129.96
129.98
132.07

You will note that some signal values were above the value of the auctioned item, and some were below the value of the item. Over a sufficiently long series of auctions, the differences between your private information signal and the value of the item will average out to zero (or very close to it). For any given auction, however, your private information signal can be above or below the value of the item. That's the nature of the random selection process generating the signals.

You will also note that V^* must always be greater than or equal to your signal value $-\,\varepsilon$. The computer calculates this for you and notes it. Further, V^* must always be less than or equal to your sample value $+\,\varepsilon$. The computer calculates this for you and notes it.

Finally, you may receive a signal value below \$25.00 (or above \$225.00). There is nothing strange about this, it just indicates V^* is close to \$25.00 (or \$225.00) relative to the size of epsilon.

8. Your signal values are strictly private information and are not to be revealed to anyone else prior to opening the bids.

You will be told the value of ε prior to bidding and it will be posted on the blackboard. However, you will not be told the value of V^* until after the bids have been posted. Finally we will post all of the signal values drawn along with the bids.

9. No one may bid less than $0.00 for the item. Nor may anyone bid more than their signal value $+\varepsilon$. Any bid in between these two values is acceptable.

Bids must be rounded to the nearest penny to be accepted.

In case of ties for the high bid, we will flip a coin to determine who will earn the item.

10. You are not to reveal your bids, or profits, nor are you to speak to any other subject while the experiment is in progress.

11. As promised, everyone will receive $4 irrespective of their earnings for participating in the experiment.

Let's summarize the main points: (1) High bidder earns the item and earns a profit = value of item − high bid price. (2) Profits will be added to your starting balance of $10.00, losses subtracted from it. Your balance at the end of experiment will be paid in cash. If balance turns negative you're no longer allowed to bid. (3) Your private information signal is randomly drawn from the interval $V^* - \varepsilon, V^* + \varepsilon$. The value of the item can never be more than your signal value $+\varepsilon$, or less than your signal value $-\varepsilon$. (4) The value of the item will always be between $25.00 and $225.00.

Are there any questions?

ADDITIONAL INSTRUCTIONS: PERIODS WITH PUBLIC INFORMATION

1. From now on bidding will be done twice during each trading period, once under each of two different information conditions. First, you will bid on the basis of your private information signals, just as you have been doing. After these bids have been made and collected, but before they are opened, you will be provided with additional information (to be described shortly) concerning the value of the item *and be asked to bid again on the commodity*. This additional information will be posted on the blackboard for everyone to see and will be referred to as a public information signal.

2. The public information signal will consist of posting on the blackboard the *lowest* of the private information signals any of you received. Note we will not reveal the bid of the player with the lowest information signal, just the signal value.

3. Note that V^* does *not* change between auctions. Your private information signals do *not* change between auctions either. However, what the public information signal does do is provide *everyone* with additional information about the possible value of V^*.

4. After both sets of bids have been collected they will be opened and the bids posted in each market and the high bid noted. We will also post the value of the item and compute profits and/or losses in the two markets as before:

PROFITS = (VALUE OF ITEM) − (HIGH BID PRICE).

Finally, to speed things up a bit we will no longer post all of the signal values drawn along with the bids.

5. However, we will only actually pay profits (or hold you accountable for losses) in one of the two markets. We will flip a coin to decide which market to pay off in. Heads we pay off in the market with private information values only, tails we pay off in the market with private and public information.

6. There is no obligation to make the same bid, or to bid differently in the two markets. This is strictly up to you to decide what to do in terms of what you think will generate the greatest profits.

Are there any questions?

REFERENCES

Baltagi, Badi H., "Pooling Cross-Sections with Unequal Time-Series Lengths," *Economics Letters*, 1986, *18*, 133–36.

Battalio, Raymond C., Kogut, Carl and Meyer, Donald J., "Individual and Market Bidding Behavior in a Vickery First Price Auction: Varying Market Size and Information," paper presented at Econometric Society Winter Meetings, 1983.

Capen, E. C., Clapp, R. V. and Campbell, W. M., "Competitive Bidding in High-Risk Situations," *Journal of Petroleum Technology*, June 1971, *23*, 641–53.

Cassing, James and Douglas, Richard W., "Implications of the Auction Mechanism in Baseball's Free Agent Draft," *Southern Economic Journal*, July 1980, *47*, 110–21.

Cox, James C., Roberson, Bruce and Smith, Vernon L., "Theory and Behavior of Single Object Auctions," in V. L. Smith, ed., *Research in Experimental Economics*, Vol. 2, Greenwich: JAI Press, 1982.

Dessauer, John P., *Book Publishing*, New York: Bowker, 1981.

Kagel, John H. et al., "First-Price, Sealed-Bid, Common Value Auctions: Some Initial Experimental Results," Center for Public Policy Discussion Paper 84–1, University of Houston, 1984.

———, "First Price Common Value Auctions: Bidder Behavior and the 'Winner's Curse'," mimeo., University of Houston, 1986.

Kagel, John H., Harstad, Ronald M. and Levin, Dan, "Information Impact and Allocation Rules in Auctions with Affiliated Private Values: A Laboratory Study," mimeo., University of Houston, 1986.

______ **and Levin, Dan,** "Individual Bidder Behavior in First-Price Private Value Auctions," *Economics Letters*, 1985, *19*, 125–28.

Leitzinger, Jeffrey J. and Stiglitz, Joseph E., "Information Externalities in Oil and Gas Leasing," *Contemporary Policy Issues*, March 1984, *5*, 44–57.

Libecap, Gary D. and Wiggins, Steven N., "The Influence of Private Failure on Regulation: The Case of Oil Field Unitization," *Journal of Political Economy*, August 1985, *93*, 690–714.

Lorenz, John and Dougherty, E. L., "Bonus Bidding and Bottom Lines: Federal Offshore Oil and Gas," SPE 12024, 58th Annual Fall Technical Conference, October 1983.

Matthews, Steven A., "Comparing Auctions for Risk Averse Buyers: A Buyer's Point of View," *Econometrica*, forthcoming 1986.

Mead, Walter, J., Moseidjord, Asbjorn and Sorensen, Philip E., "The Rate of Return Earned by Leases Under Cash Bonus Bidding in OCS Oil and Gas Leases," *Energy Journal*, October 1983, *4*, 37–52.

______, ______, **and** ______, "Competitive Bidding Under Asymmetrical Information: Behavior and Performance in Gulf of Mexico Drainage Lease Sales, 1959–1969," *Review of Economics and Statistics*, August 1984, *66*, 505–08.

Milgrom, Paul R. and Weber, Robert J., "A Theory of Auctions and Competitive Bidding," *Econometrica*, September 1982, *50*, 1089–122.

Plott, Charles R., "Industrial Organization Theory and Experimental Economics," *Journal of Economic Literature*, December 1982, *20*, 1485–527.

Roll, Richard, "The Hubris Hypothesis of Corporate Takeovers," *Journal of Business*, April 1986, *59*, 197–216.

Smith, Vernon L., "Microeconomic Systems as an Experimental Science," *American Economic Review*, December 1982, *72*, 923–55.

Weverbergh, M., "Competitive Bidding With Asymmetric Information Reanalyzed," *Management Science*, March 1979, *25*, 291–94.

Wilson, Robert, (1975a) "Comment on David Hughart, Informational Assymetry, Bidding Strategies and the Marketing of Offshore Petroleum Leases," mimeo., Stanford University, 1975.

______, (1975b) "On the Incentive for Information Acquistion in Competitive Bidding with Asymmetrical Information," mimeo., Stanford University, 1975.

______, "A Bidding Model of Perfect Competition," *Review of Economic Studies*, October 1977, *44*, 511–18.

______, "The Basic Model of Competitive Bidding," mimeo., Stanford University, 1981.

Winer, Ben J., *Statistical Principles in Experimental Design*, 2nd ed., New York: McGraw-Hill, 1971.

The Logic of Vertical Restraints

By PATRICK REY AND JEAN TIROLE*

This paper shows that direct (product) competition acts as a tournament between retailers when informational problems (or transaction costs) prevent the manufacturer from using contracts based on their relative performances. Anticompetitive restraints such as exclusive territories and resale price maintenance (which, we show, are not necessarily good substitutes) may or may not be privately desirable. It also shows that privately desirable anticompetitive restraints may not be socially desirable.

Upstream and downstream firms often do not trade intermediate goods through a simple linear price mechanism. Their relationship also involves vertical restraints such as resale price maintenance (*RPM*), exclusive territories (*ET*), franchises and exclusive dealing, to name a few. Vertical restraints constitute a challenge to both the theorist and the applied economist. They represent one of the most obvious applications of agency theory, with relatively explicit contracting terms. They are of considerable interest to the antitrust practitioner, as witnessed by the high number of cases related to such restraints. Furthermore, some disagreement remains as to what motivates these restraints, and even more disagreement relates to their welfare consequences (see F. M. Scherer, 1980; Roger Blair and David Kaserman, 1983).

This paper's tenet is that the private and social desirability of such restraints ought to be studied by first analyzing the delegation problem underlying them. This is indeed the route taken by the recent theoretical literature on this topic (Avinash Dixit, 1983; Nancy Gallini and Ralph Winter, 1983; G. Frank Mathewson and Winter, 1983a,b,c; 1984). In this literature, the upstream unit (manufacturer) sells a good to downstream units (retailers). The manufacturer can control/observe some variables such as the quantity bought by each retailer and possibly the retailers' prices, their areas of distribution, etc. However, it either cannot, or is not legally allowed to, control some other variables such as the number of retailers, their locations, or their selling efforts. This lack of control of some variables gives rise to undesirable horizontal and vertical externalities. For instance, the lack of control of selling expenses may give rise to underprovision of effort and free riding (Lester Telser, 1960); or imperfect control of retail prices gives rise to the celebrated double marginalization (Joseph Spengler, 1950). The main merit of the above-mentioned literature is to show how simple vertical restraints may suppress these externalities. To cite only two well-developed examples, a franchise (or *RPM*) can solve the double marginalization problem; while *ET* or *RPM* (together with a franchise) can eliminate promotional externalities. Mathewson and Winter (1983a, 1984) offer a fairly comprehensive treatment of this type of control.

This literature derives some interesting and intuitive explanations of vertical restraints.[1] It, however, has the drawback of not exploring the roots of the delegation problem. Indeed, each agent (retailer) makes its decisions having the same full information as the principal (manufacturer) about his certain environment. Thus the manufacturer's only concern is to make sure that the retailers take some given actions. The adaptation to the environment through delegation is ignored because of the absence of uncertainty. In this literature, the manufacturer generally uses all available vertical restraints (for instance, if he can monitor the final price, he imposes resale price maintenance, since, at worst, he can duplicate the competitive price outcome). Indeed, a theme of this literature is the search for a set of "minimally sufficient vertical restraints," that is, restraints that enable the manufacturer to realize the first-best or integrated profit.

---

*Ecole Nationale de la Statistique et de l'Administration Economique (ENSAE), Paris, 92240 Malakoff, France, and Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139, respectively. We are grateful to Paul Joskow, Ralph Winter, a referee, and participants of the Microworkshop at MIT and the conference on distribution organized by CESCOM in Angera (Italy), where a first draft of this paper was presented. This research was supported in part by the National Science Foundation and the Sloan Foundation.

[1] For a review of the methodology of this literature, see our earlier paper (1986).

In this paper we consider retailers' informational superiority about the environment as well. Given this informational assumption, we focus on the private and social desirability of anticompetitive restraints such as *RPM* and *ET*. We compare them with the policy that allows competition between the retailers. We find that:

(a) vertical restraints may or may not be desirable from the manufacturer's point of view, that is, may or may not be part of an optimal contract. Thus, the manufacturer may decide not to use informationally feasible restraints. Indeed, a central theme of the paper is that direct (product) competition acts as a tournament between the retailers when informational problems prevent the manufacturer from using contracts based on their relative performances.

(b) in contrast to conventional wisdom, *RPM* and *ET* are not necessarily substitutes;

(c) the comparison between the different control policies depends much on the type of uncertainty; and

(d) vertical restraints may both be privately desirable and decrease aggregate surplus, and therefore should not be legal per se.

Section I sets up the model. Section II computes the manufacturer's profit and social welfare for competition, *ET* and *RPM*, under "market uncertainty" about cost and demand and retailers' extreme risk aversion. It compares the results for this case, and discusses the factors influencing the private desirability of restraints. Section III makes the polar assumption of retailers' risk neutrality, while Section IV treats the case of "idiosyncratic uncertainty." Section V summarizes the welfare aspects and explains the likely bias of the social planner towards competition. Section VI generalizes our analysis to arbitrary concave utility functions for the retailers. It studies the "purely competitive" framework—market uncertainty and undifferentiated retailers—and shows that the manufacturer always prefers competition to *RPM*, and that the difference in the manufacturer's profit between competition and *ET* grows with the retailers' Arrow-Pratt index of absolute risk aversion; the manufacturer prefers competition (*ET*) if the retailers are very risk averse (risk neutral); it also shows that aggregate welfare is always higher under competition than under an anticompetitive restraint (*ET* or *RPM*). Lastly, Section VII draws the main conclusions of our analysis and suggests some extensions.

I. The Model

A. *Description*

Let us set the basic model. Some straightforward variants will be introduced later to study *ET* and *RPM*. A manufacturer produces a single product at constant unit cost c. He supplies this product to a large number of "independent markets." (One can think of geographical markets.) A market, the entity we focus on, stands for a set of customers located on a segment of length one. They are uniformly distributed along the segment with unit density. Consumers are indexed by their distance y to the left end of the segment.

We will not be concerned with the number of retailers or their location (these features could be incorporated into the model). We simply assume that there are two retailing sites located at the two extremes of the market. Each site is occupied by a retailer, who is chosen by the manufacturer. There is a competitive supply of (identical) potential retailers.

A consumer located at y pays a transportation cost ty (respectively, $t(1-y)$) per unit of product purchased from the left-end retailer (respectively, right-end retailer). For instance, if he picks the left-end retailer and the latter charges a retail price q, the real price paid by the consumer is $\{q+ty\}$. We assume that all consumers have the same linear demand curve. Let $x=d-(q+ty)$ denote the quantity bought by a consumer located at y, assuming he buys from the left-end retailer.

A retailer has distribution cost γ per unit of sale. He chooses the retail price of the product. The two retailers are Bertrand competitors (as shown in Section V, the absence of collusion between the retailers is crucial to some of our results).

We assume for the moment that the manufacturer can impose only two-part tariffs on his retailers. So if x is the quantity sold to a retailer, the latter pays $\{A+px\}$ to the manufacturer, where p is the wholesale price and A is a franchise fee. The next section argues that under reasonable informational asymmetries between the manufacturer and the retailers, two-part tariffs are indeed optimal for the manufacturer. For the moment, let us simply justify two-part tariffs by their common use in real-world vertical arrangements, or else by the prohibition by some state regulations that the same product be sold to different dealers at different marginal prices.

Let us now describe the uncertainty about the environment. As a first step, we consider only uncertainty about each market, and not about each retailer. This assumption implies that the two retailers in a given market are still symmetrical once the uncertainty is resolved. This market uncertainty is market specific in that the random variables in different markets are uncorrelated. In Section IV, we will consider the case of independent or idiosyncratic uncertainty affecting each retailer.

We simultaneously consider two types of market-specific uncertainty:

1) *Demand Uncertainty*: The demand parameter d is distributed on $[\underline{d},\bar{d}]$. The parameter d is not known at the time the contracts with the retailers are signed.

2) *Retail Cost Uncertainty*: The distribution cost γ is distributed on $[\underline{\gamma},\bar{\gamma}]$. The parameter γ is not known at the time the contracts are signed. We will assume that $\underline{d}>c+\bar{\gamma}$, and that d and γ are independent.

The uncertainty (d and γ) is not observed by the manufacturer. It is observed by the two retailers after their contract is signed, but before they take their pricing decisions. We thus emphasize the adjustment of decision making to the environment.

We posit that the manufacturer is risk neutral, which can be justified by our assumption that he (or she) supplies a large number of independent markets. We first assume that a potential retailer is infinitely risk averse; and that he requires that his profit in the worst possible outcome (the only one he cares about *ex ante*) be nonnegative. In Section III, we consider the polar case in which potential retailers are risk neutral and Section VI shows that our conclusions extend to the general case. Lastly, we assume that retailers must supply demand at the quoted price (this assumption simplifies the analysis, but is not restrictive).

In most of the paper we give the derivations only for the case of zero transportation cost ($t=0$). We call this the case of *undifferentiated retailers*. Most of the intuition can be obtained from this case. Our construction of a more general Hotelling differentiation model serves two purposes, however. First, it helps motivate the notion of exclusive territories introduced later. Second, when alternative policies are tied in the undifferentiated retailers case, nonzero transportation costs break the tie (see Section II).

B. *Two-Part Tariffs as Optimal Contracts*

Let us justify two-part tariffs from (more primitive) informational assumptions. To this purpose, let us make the following assumptions: (a) The manufacturer can observe only the amount of product supplied to a retailer and whether or not the retailer carries his product. He can observe neither the actual quantity sold by the retailer and the latter's profit, nor for the moment the retail price and the area of distribution.[2] (b) The manufacturer is not allowed to refuse to deal (*ex post*) with a retailer (see fn. 5 below).

Under these informational assumptions, if the manufacturer tries to price discriminate according to the quantity bought by retailers (i.e., does not charge a constant marginal price), and if, as we assumed, he serves a high number of independent markets, the retailers may set up a secondary market for the good ("bootleg"), provided that their transportation costs are low enough.[3] In the extreme case in which the secondary market for the product is perfect, the manufacturer cannot make money by not charging a constant marginal price (equal to the secondary market price). He can, however, impose a franchise fee as he is informed that the retailer carries his product. Note in particular that quantity forcing—an instance of discrimination—is not effective here.

Remark: The reader familiar with the tournament literature may presume (perfect) correlation between the two retailers' environments may enable the manufacturer to extract more information and use more efficient contracts. Remember, however, that the final sales cannot be observed. Furthermore, the existence of a secondary market prevents the quantities purchased by the retailers from conveying any information. Thus, any information that the manufacturer obtains about the state of nature must come from the retailers' announcing it. However, these announcements are "costless"; and it is easily seen that the set of equilibria and payoffs of an announcement game between retailers organized by the manufacturer is independent of the realized state of nature.[4] Hence, under the reasonable assumption that the retailers consistently coordinate on the same announcement equilibrium (i.e., whatever the state of nature), the manufacturer does not gain by designing such an announcement game.[5] So the two-part tariff is indeed optimal. The economic intuition behind this reasoning is that the retailers can always coordinate to announce that they are facing an adverse environment.

II. Market Uncertainty and Extreme Retailer Risk Aversion

A. *Competition*

As discussed above, the manufacturer, if he cannot control retail prices and areas of distribution, offers a two-part tariff to his retailers. A contract is, therefore, the choice of a franchise fee A and a wholesale price p.

---

[2] The nonobservability of a retailer price has two alternative interpretations. In the first, the retail price itself is not observable, or at least not verifiable by a court, because the retailer can give hidden price discounts to his customers. In the second, the retail price q is observed. But the retailer supplies some services (effort) that are not verifiable. Suppose that these services have monetary value s per unit of good for the consumer. Also suppose that the cost of supplying these services is $\Phi(s) = s$ per unit of good. Then, by defining $\tilde{q} = q - s$, everything (from the point of view of the retailer and consumers) is as if the retailer chose the net price $\tilde{q}$ and this net price was not verifiable by a court. Thus our model can alternatively be interpreted as one of retailers' incentives. It can easily be shown that our models of competition and exclusive territories below hold under this alternative interpretation. In particular, it is pointless to construct more complex contracts than the ones studied here.

[3] It may seem strange that in a model in which transportation costs may play a role within a market, we neglect transportation costs between markets. Three justifications can be given for this. First, our results hold for zero or negligible consumers' transportation costs within a market. Second, retailers' transportation costs per unit of distance and product may be much lower than those of consumers. Third, the retailers can be differentiated by other attributes than location.

[4] At least in the two polar cases of risk aversion or in the case of constant absolute risk aversion (this statement also holds for deterministic equilibria for any utility function). The possibility of multiple Nash equilibria in a multi-agent context was noted by Dilip Mookherjee (1984). Its implications are particularly severe in our model.

[5] Our assumption that the manufacturer cannot refuse to deal with a retailer rules out the *ex post* auctioning off of the market between the two retailers under contract. This assumption is a crude but simple way to formalize the idea that the manufacturer needs the two retailers. It can easily be derived endogenously from the existence of private goodwill (or sufficient differentiation) or of increasing marginal distribution costs. (In a more general model, the number of retailers is in particular determined by the efficient scale.)

We restrict ourselves to symmetric contracts, which involves no loss of generality.[6] We compute the equilibrium behaviors, profits, and welfare given that the two retailers compete through prices.

Initially, we consider the case of *undifferentiated retailers*. From the absence of differentiation, the two retailers act as Bertrand competitors and charge the same retail price in equilibrium: $q = p + \gamma$. Furthermore, whatever the state of nature, the retailers make no *ex post* profit. The franchise fee must therefore equal zero.

Since each consumer's demand is $\{d - p - \gamma\}$, the manufacturer's expected profit Π^C is given by

$$\Pi^C = \max_p E\{(d-p-\gamma)(p-c)\}$$
$$= \max_p (d^e - p - \gamma^e)(p-c),$$

where $d^e \equiv E_d d$ is the expectation of the demand parameter and $\gamma^e \equiv E_\gamma \gamma$ the expectation of the retail cost. This yields, with obvious notation,

$$A^C = 0, \tag{1}$$

$$p^C = \tfrac{1}{2}(d^e + c - \gamma^e), \tag{2}$$

$$q^C = \tfrac{1}{2}(d^e + c - \gamma^e) + \gamma, \tag{3}$$

$$\Pi^C = \tfrac{1}{4}(d^e - c - \gamma^e)^2. \tag{4}$$

Defining aggregate welfare W^C as the expected sum of consumers' and manufacturer's surpluses (by assumption the retailers have no surplus), we get

$$W^C = E\left(\tfrac{1}{2}(d-q)^2\right) + \Pi^C$$

or

$$W^C = \tfrac{3}{8}(d^e - c - \gamma^e)^2 + \tfrac{1}{2}\sigma_d^2 + \tfrac{1}{2}\sigma_\gamma^2, \tag{5}$$

where $\sigma_d^2 \equiv E_d(d - d^e)^2$ is the variance of d and $\sigma_\gamma^2 \equiv E_\gamma(\gamma - \gamma^e)^2$ is the variance of γ.

Note that for both types of uncertainty the retailers' surplus is equal to zero not only *ex ante*, but also *ex post*. So, given the wholesale price and the retailers' behavior, the franchise fee does an extremely good job at capturing the retailers' surplus. Note also that the consumer price q^C adjusts fully to retail cost variations and does not adjust to demand conditions, so that the final demand $\{x = d - q\}$ fully responds to both types of shocks. These properties will prove to be important in the following comparisons.

B. *Exclusive Territories*

Let us now slightly modify our informational assumptions by assuming that the manufacturer can also observe the areas of distribution of the two retailers. In particular, he can assign the left-end half of the market to the left-end retailer, and the right-end one to the right-end retailer.[7] This practice, the exclusive territory arrangement (*ET*), is common in franchise contracts. Each retailer buys the product according to the two-part tariff as long as he does not infringe on the other retailer's territory; otherwise he is heavily fined. We maintain our assumption that the manufacturer cannot control the retail price, so that the optimal contract is either a two-part tariff contract with competition or a two-part tariff contract with *ET*.

Let us now compute the optimal two-part tariff, profit, and welfare under *ET*. Each retailer is a monopoly in his territory. Given a wholesale price p and realizations of demand d and cost γ, he maximizes $\{\frac{1}{2}(d-q)(q-p-\gamma)\}$ over his resale price q. So $q = \frac{1}{2}(d+p+\gamma)$; and the retailer's *ex post* profit is $\frac{1}{8}(d-p-\gamma)^2$. As he is extremely risk averse, the franchise A must give him a nonnegative profit in the worst state of nature. Thus, $A = \frac{1}{8}(\underline{d} - p - \bar{\gamma})^2$. The total franchise fee received by the manufacturer in the market is $2A = \frac{1}{4}(\underline{d} - p - \bar{\gamma})^2$.

---

[6] As explained earlier, p must be the same for the two retailers. Furthermore, A governs only the retailers' acceptance decision, but not their *ex post* behavior.

[7] More complex territorial assignments turn out to be unprofitable.

The manufacturer's optimization problem can be written

$$\Pi^{ET} = \max_{p} \left\{ \tfrac{1}{4}(\underline{d} - p - \bar{\gamma})^2 \right.$$
$$\left. + E\big(d - \tfrac{1}{2}(d + p + \gamma)\big)(p - c) \right\}.$$

This yields

$$A^{ET} = \tfrac{1}{8}(\underline{d} - p - \bar{\gamma})^2, \tag{6}$$

$$p^{ET} = c + (d^e - \underline{d}) + (\bar{\gamma} - \gamma^e), \tag{7}$$

$$q^{ET} = \tfrac{1}{2}\big(d + c + \gamma + (d^e - \underline{d}) + (\bar{\gamma} - \gamma^e)\big), \tag{8}$$

$$\Pi^{ET} = \tfrac{1}{4}(\underline{d} - c - \bar{\gamma})^2 + \tfrac{1}{4}[(d^e - \underline{d}) + (\bar{\gamma} - \gamma^e)]^2, \tag{9}$$

$$W^{ET} = \tfrac{3}{8}[\underline{d} - c - \bar{\gamma}]^2 + \tfrac{1}{4}[(d^e - \underline{d}) + (\bar{\gamma} - \gamma^e)]^2 + \tfrac{1}{8}\sigma_d^2 + \tfrac{1}{8}\sigma_\gamma^2. \tag{10}$$

Note that under demand uncertainty, the wholesale price exceeds marginal cost ($p > c$). They are equal only in the certainty case ($d^e = \underline{d}$ and $\gamma^e = \bar{\gamma}$). Under uncertainty, a wholesale price close to marginal cost is still desirable to avoid double marginalization. However, the uncertainty faced by a retailer under *ET* decreases with the wholesale price: a higher wholesale price means a lower retailer's profit margin and therefore a lower sensitivity of retailer's profit to demand and cost variations. Thus, increasing the wholesale price above marginal cost has desirable insurance properties, in that it transfers risk bearing from the retailers to the risk-averse manufacturer.

C. *Resale Price Maintenance*

Let us now assume that, in addition to the quantity bought by a retailer and the fact that the latter carries the product, the manufacturer also observes the retail price (see fn. 2 for the limitations of such an assumption). A feasible policy then consists in fixing the retailer's retail price, together with a two-part tariff for his purchase of the product. We call this policy resale price maintenance (*RPM*). In the presence of price observability, there may exist better incentive schemes than competition (two-part tariff) and *RPM* (two-part tariff plus resale price fixing).[8] The manufacturer could more generally charge a two-part tariff in which the franchise fee would depend on the price chosen by the retailer (and possibly on the price chosen by his competitor), if such a contract can be enforced. Unless the law prohibits discrimination between retailers (which it does in France and the United States), focusing on competition and *RPM* involves some loss of generality. But these two arrangements are interesting in their own right, as they are both simple and frequently used. Let us also mention that, in this particular model, price ceilings and price floors can be shown not to be optimal for the manufacturer (they are dominated either by *RPM* or by competition).

Let us now solve for the best *RPM* contract. We assume without loss of generality that the same contract is offered to all retailers.

Under *RPM*, a retailer is rather passive. He charges the prescribed price q and obtains half of the market. His *ex post* profit is $\{\frac{1}{2}(d - q)(q - p - \gamma)\}$ in state of demand d and cost γ. Thus, the franchise must be $A = \frac{1}{2}(\underline{d} - q)(q - p - \bar{\gamma})$ (for $q \geq p + \bar{\gamma}$; for $q < p + \bar{\gamma}$, $A = \frac{1}{2}(\bar{d} - q)(q - p - \bar{\gamma})$: it is easily shown that one can restrict oneself to the situation $q \geq p + \bar{\gamma}$).

Resale price maintenance (*RPM*) adds both an extra control variable plus an extra constraint (that the price be constant across states of nature) to the manufacturer's maxi-

---

[8] Note that, if the area of distribution is also observable, *ET* plus *RPM* is equivalent to *RPM*. This property may not hold in different models (see Mathewson-Winter, 1984).

TABLE 1—UNDIFFERENTIATED RETAILERS, MARKET UNCERTAINTY, AND EXTREME RISK AVERSION

	Retail Price	Sensitivity of Retail Price to Uncertainty	Private Desirability of Restraints	Social Desirability of Restraints
Demand Uncertainty	$\hat{q}^{ET} > \hat{q}^{C} = \hat{q}^{RPM}$	$d\hat{q}^{ET}/dd = .5$		
$(\gamma^e = \bar{\gamma},\ \sigma_\gamma^2 = 0)$	for all d	$d\hat{q}^{c}/dd = 0$	$\hat{\Pi}^C = \hat{\Pi}^{RPM} > \hat{\Pi}^{ET}$	$\hat{W}^C = \hat{W}^{RPM} > \hat{W}^{ET}$
		$d\hat{q}^{RPM}/dd = 0$		
Cost Uncertainty	$E_\gamma(\tilde{q}^{RPM}) =$	$d\tilde{q}^{C}/d\gamma = 1$		
$(d^e = \underline{d},\ \sigma_d^2 = 0)$	$E_\gamma(\tilde{q}^{ET}) > E_\gamma(\tilde{q}^{C})$	$d\tilde{q}^{ET}/d\gamma = .5$	$\tilde{\Pi}^C > \tilde{\Pi}^{ET} > \tilde{\Pi}^{RPM}$	$\tilde{W}^C > \tilde{W}^{ET} > \tilde{W}^{RPM}$
		$d\tilde{q}^{RPM}/d\gamma = 0$		

mization problem, which becomes:

$$\Pi^{RPM} = \max_{\{(p,q)|q \geq p+\bar{\gamma}\}} \{(\underline{d} - q)(q - p - \bar{\gamma}) + E_d(d - q)(p - c)\}.$$

We obtain

$$(11)\quad A^{RPM} = 0,$$

$$(12)\quad p^{RPM} = \tfrac{1}{2}(d^e + c - \bar{\gamma}),$$

$$(13)\quad q^{RPM} = \tfrac{1}{2}(d^e + c + \bar{\gamma}),$$

$$(14)\quad \Pi^{RPM} = \tfrac{1}{4}(d^e - c - \bar{\gamma})^2,$$

$$(15)\quad W^{RPM} = \tfrac{3}{8}(d^e - c - \bar{\gamma})^2 + \tfrac{1}{2}\sigma_d^2.$$

D. *Comparisons and Economic Intuition*

We first compare the results obtained so far for undifferentiated retailers. Straightforward computations lead to Table 1 ("hats" are used to denote demand uncertainty, and tildes are used to denote cost uncertainty). We see that, under demand uncertainty, the manufacturer is indifferent between competition and *RPM* (assuming that he can observe the retailers' prices). To break this tie, let us consider the case of differentiated retailers ($t > 0$). The computations are somewhat tedious. We simply give here the main results for *slightly* differentiated retailers (t small, but positive):

$$(16)\quad \hat{\Pi}^{RPM} > \hat{\Pi}^{C};$$

$$(17)\quad \hat{q}^{RPM} \simeq \hat{q}^{C} \text{ (to the second order in } t).$$

Note that (16) and (17) imply that the social planner also prefers *RPM* to competition.

These results, Table 1, and continuity with respect to the differentiation parameter (which can be checked) lead to the following propositions:

PROPOSITION 1: *If there is no uncertainty* ($\sigma_d^2 = \sigma_\gamma^2 = 0$), *competition, ET and RPM are equivalent, both from the manufacturer's and from the social welfare points of view. The retail price and the manufacturer's profit are equal to the vertically integrated monopoly price and profit.*

PROPOSITION 2: *Assume that retailers are slightly differentiated and extremely risk averse. Under uncertainty, competition, ET and RPM are not equivalent: under demand uncertainty, the manufacturer prefers RPM to competition, and competition to ET. Under cost uncertainty he prefers competition to ET, and ET to RPM. The social planner does not object to these rankings.*

Proposition 1 is familiar from the literature: for our model and under certainty, a two-part tariff, *RPM* and *ET* are all sufficient control mechanisms in the sense of Mathewson-Winter (1984), in that they allow the manufacturer to realize the vertically integrated profit.

Proposition 2 summarizes the basic results for market demand and cost uncertainty. Let us now give the economic intuition behind the *manufacturer's* choices between competition and vertical restraints (welfare aspects will be studied in Section V).

The manufacturer has two objectives:

1) To ensure an optimal exploitation of monopoly power by the vertical structure; and
2) To provide adequate insurance to his retailers.

These two objectives, which aim at not choking final demand and at extracting a high franchise fee from the retailers, conflict, as we shall see shortly.

The *optimal exploitation of monopoly power* has been the focus of the literature on vertical restraints. The manufacturer would like the retail price to be as close as possible to the monopoly price (the price that would be charged by the vertically integrated industry), and this *for all states of nature*. With linear demand this monopoly price is $q^m = (d + c + \gamma)/2$. It is responsive, although not fully, to demand and retail cost shocks. The degree of responsiveness will be the focus of our discussion, as the manufacturer can always control the *average* retail price through the wholesale price or *RPM*.

Consider first exclusive territories. Under this restraint, each retailer has an unconstrained monopoly on his territory, and therefore chooses the correct response to a demand or retail cost shock. The one difference with the vertically integrated industry is that his fictitious production cost is the wholesale price p instead of the marginal cost c. Under linear demand, he charges $q = (d + p + \gamma)/2$. So we conclude that exclusive territories ensure a perfect use of decentralized information. To obtain a perfect coincidence of the retailer's and the vertical structure's objectives, the manufacturer must avoid a pricing distortion at the wholesale level: $p = c$. This is indeed what he does under certainty (see equation (7)); under uncertainty, however, this policy conflicts with insurance as we will see. Second, if the manufacturer imposes *RPM*, the retail price exhibits no responsiveness to demand and retail cost shocks. (Price ceilings and floors do allow some use of decentralized information, but are nonoptimal in this model.)

Lastly let us consider competition. When the retailers are undifferentiated, the retail price is entirely "cost determined": $q = p + \gamma$. As for *RPM*, this price is insensitive to demand shocks. Contrary to *RPM*, it reacts to retail cost shocks; it even overreacts relatively to the vertically integrated structure. When retailers are differentiated, the retail price reacts more to demand shocks and less to cost shocks, as the competition weakens and the structure becomes closer to the *ET* arrangement. From this study, we conclude that *ET* uses decentralized information in a more efficient way than *RPM* and competition; it is superior in terms of exploitation of monopoly power, as the retail price can be fine-tuned to the retailers' environment.

We now study the *insurance objective*. Let us first consider competition. If the retailers are undifferentiated, Bertrand competition ensures that their profit is not sensitive to uncertainty, so competition provides perfect insurance to the retailers. The profit of differentiated retailers does fluctuate under uncertainty, but the insurance properties of competition remain fairly good. *RPM* has terrible insurance properties under cost uncertainty: the retailers then bear the whole risk as they must fully adjust their profit margin to cost shocks. Under demand uncertainty, however, *RPM* gives perfect insurance. The profit margin is independent of the state of nature, and the wholesale price can be chosen so that this profit margin be zero, in which case the retailers' profit is independent of demand (see equations (12) and (13)). This perfect insurance is the reason why the manufacturer prefers *RPM* to competition under demand uncertainty and differentiated retailers. Lastly, consider exclusive territories. The retailers adjust their price to the shocks, but cannot avoid substantial fluctuations in their profits. To share the retailers' risk somewhat, the manufacturer is forced to raise the wholesale price above the marginal cost, thus introducing a double marginalization.

To summarize: *ET* makes a better use of decentralized information than competition or *RPM*. Competition has very good insurance properties under both types of uncertainty. *RPM* gives perfect insurance under demand uncertainty, but lets the retailers bear the whole risk under cost uncertainty. *ET* has mediocre insurance properties.

Let us demonstrate that *ET* is not optimal under market uncertainty and infinite risk aversion, by noticing that under an appropriate competitive scheme the manufacturer can always make more than the *ET* profit by (a) requiring no franchise, and (b) charging a wholesale price equal to the *ET* retail price in the worst state of nature ($\underline{d}$ and $\bar{\gamma}$) minus the highest possible distribution cost. The retailers clearly accept such a contract. If $q^m(p+\gamma, d)$ denotes the monopoly price in state of demand d with total input price $(p+\gamma)$, the manufacturer's profit under *ET* is

$$\begin{aligned}&[\underline{d} - q^m(p^{ET}+\bar{\gamma}, \underline{d})]\\ &\qquad\times[q^m(p^{ET}+\bar{\gamma}, \underline{d}) - p^{ET} - \bar{\gamma}]\\ &\qquad+E[d - q^m(p^{ET}+\gamma, d)][p^{ET} - c],\end{aligned}$$

and, under the proposed competitive arrangement, it is

$$\begin{aligned}&E[d - (q^m(p^{ET}+\bar{\gamma}, \underline{d}) - \bar{\gamma} + \gamma)]\\ &\qquad\times[q^m(p^{ET}+\bar{\gamma}, \underline{d}) - \bar{\gamma} - c].\end{aligned}$$

Using successively the facts that for all d, $\underline{d} - q^m(p^{ET} + \bar{\gamma}, \underline{d}) \leq d - q^m(p^{ET} + \bar{\gamma}, d)$ (final demand is higher in higher states of demand) and that for all γ and d, $q^m(p^{ET} + \bar{\gamma}, d) \geq q^m(p^{ET} + \gamma, d)$ (the retail price increases with the retail cost under monopoly), we see that the manufacturer prefers the proposed competitive arrangement to *ET*. The basic idea of this proof is that by imposing competition, the manufacturer can do as well in the worst state of nature (as in the certainty case); and he keeps the retail price down and thus demand up, in better states of nature. This reasoning holds for undifferentiated retailers. Under differentiation, a similar argument can be made that *ET* is dominated by *RPM*.

Lastly, under demand uncertainty and differentiated retailers, the manufacturer does a little bit better by using *RPM* rather than competition. In the competition case differentiation introduces a double marginalization which *RPM* helps suppress.

The main conclusions of this section are that competition has good insurance properties and does relatively well (it is only dominated by *RPM* in the case of demand uncertainty and differentiated retailers). *RPM* and *ET* are far from being substitutes. *RPM* is preferred in case of demand uncertainty, while *ET* scores better for cost uncertainty.

III. Market Uncertainty and Retailer Risk Neutrality

We now show that the degree of risk aversion is crucial to the previous rankings. We still assume that the manufacturer is risk neutral. But we now consider the polar case in which the retailers are also *risk neutral*. For simplicity we study undifferentiated retailers.

Under *competition*, the retailers make a zero variable profit whatever the state of nature. So risk aversion plays no role and $(A^C, p^C, q^C, \pi^C, W^C)$ are still given by equations (1) through (5).

Under *ET*, in state of nature (d, γ), a retailer charges $q = \frac{1}{2}(d + p + \gamma)$ and makes profit $\frac{1}{8}(d - p - \gamma)^2$. The essential difference with the extreme risk-aversion case is that now the manufacturer can recover this profit in expectation, and not only in the worst state of nature. So

$$A^{ET} = \tfrac{1}{8}E(d - p - \gamma)^2. \tag{18}$$

Maximizing the manufacturer's profit over the wholesale price gives

$$p^{ET} = c \tag{19}$$

$$q^{ET} = \tfrac{1}{2}(d + c + \gamma) \tag{20}$$

$$\Pi^{ET} = \tfrac{1}{4}(d^e - c - \gamma^e)^2 + \tfrac{1}{4}\sigma_d^2 + \tfrac{1}{4}\sigma_\gamma^2 \tag{21}$$

$$W^{ET} = \tfrac{3}{8}(d^e - c - \gamma^e)^2 + \tfrac{3}{8}\sigma_d^2 + \tfrac{3}{8}\sigma_\gamma^2. \tag{22}$$

Notice that the manufacturer avoids the double distortion by charging a wholesale price equal to his marginal cost.

Lastly, under *RPM*, one can easily show that

$$q^{RPM} = \tfrac{1}{2}(d^e + \gamma^e + c) \tag{23}$$

$$\Pi^{RPM} = \tfrac{1}{4}(d^e - \gamma^e - c)^2 \tag{24}$$

$$W^{RPM} = \tfrac{3}{8}(d^e - \gamma^e - c)^2 + \tfrac{1}{2}\sigma_d^2 \tag{25}$$

(A and p are indeterminate).

PROPOSITION 3: *Assume that the retailers are risk neutral and nondifferentiated. Under either demand or cost uncertainty,* (a) *the manufacturer prefers ET to competition and RPM;* (b) *however, the social planner objects to this ranking as he prefers competition to ET (and in general to RPM as well).*

Let us now interpret and draw the implications of Proposition 3(a) (leaving 3(b) for Section V). We note in Section II that *ET* makes better use of decentralized information while competition is more efficient at providing insurance. We showed that, under extreme risk aversion, the second effect dominates the first so that the manufacturer prefers competition. Under risk neutrality, however, the second effect disappears, and therefore, *ET* is preferred. Actually, *ET* even enables the manufacturer to realize the integrated profit by making each retailer a *residual claimant* for the profits of the vertical structure.

The other interesting fact is that the social planner does favor competition. This will be explained in Section V.

Remark: As in Section II, we may wonder whether we can break the ties by introducing a differentiation between the retailers. It turns out that, under demand uncertainty and retailers' risk neutrality, competition and *RPM* are always equivalent from both the private and the social points of view: Fixing a wholesale price is equivalent to fixing the retail price, under competition and demand uncertainty. So the vertical structure can realize the same aggregate profit with the two arrangements. Since the manufacturer can always capture the retailers' expected profit under risk neutrality, he also realizes the same profit with the two arrangements.

By contrast, the equivalence between competition and *RPM* under cost uncertainty and retailers' risk neutrality is an artifact of the linearity of demand (for instance, for a convex demand curve $\tilde{\Pi}^C > \tilde{\Pi}^{RPM}$).

IV. Idiosyncratic Uncertainty

Until now we have considered only market uncertainty. We can also study the polar case in which the two retailers face (independent) idiosyncratic uncertainty. Here we naturally focus on cost uncertainty, as a demand shock will most certainly affect interdependent retailers simultaneously.

We assume that retailers are not differentiated and that each retailer knows his competitor's cost as well as his own cost when he chooses a retail price.[9] For simplicity, we also assume that each firm's distribution cost is distributed according to an exponential distribution $F(\gamma) = ((\gamma - \underline{\gamma})/\bar{\gamma} - \underline{\gamma}))^{\alpha}$, where $\alpha > 0$ ($\alpha = 1$ gives the uniform distribution) on $[\underline{\gamma}, \bar{\gamma}]$; and that $(\bar{\gamma} - \underline{\gamma})$ is "not too high," so that under competition, a retailer with cost γ_1 charges $(p + \gamma_2)$ when his competitor has cost $\gamma_2 > \gamma_1$.

It can be shown that (see the Appendix):

With infinitely risk-averse (respectively risk-neutral) retailers, both the manufacturer and the social planner prefer competition to *ET* (respectively, *ET* to competition).

RPM is always dominated from both points of view (as in the market uncertainty case).

The analysis also suggests that conflicts between social and private desirabilities will occur for intermediate degrees of risk aversion.

In our model, *RPM* and *ET* insulate retailers from intrabrand competition. Thus, the degree of correlation between the re-

---

[9] He may have inside information, for instance. However, we maintain the assumption that the manufacturer cannot make the contracts contingent on costs (i.e., costs are not verifiable by a court). Had we assumed that the retailers do not observe their competitor's cost, we would have used the Bayesian Nash equilibrium concept.

TABLE 2—WELFARE COMPARISONS

	Infinite Risk Aversion with Market Uncertainty	Risk Neutrality with Market Uncertainty	Idiosyncratic Uncertainty
Demand Uncertainty	$\hat{\Pi}^C = \hat{\Pi}^{RPM} > \hat{\Pi}^{ET}$	$\hat{\Pi}^{ET} > \hat{\Pi}^C = \hat{\Pi}^{RPM}$	
	$\hat{W}^C = \hat{W}^{RPM} > \hat{W}^{ET}$	$\hat{W}^C = \hat{W}^{RPM} > \hat{W}^{ET}$	
Retail Cost Uncertainty	$\tilde{\Pi}^C > \tilde{\Pi}^{ET} > \tilde{\Pi}^{RPM}$	$\tilde{\Pi}^{ET} > \tilde{\Pi}^C = \tilde{\Pi}^{RPM}$	$\tilde{\Pi}^C \gtrless \tilde{\Pi}^{ET}$ (*RPM* dominated by both)
	$\tilde{W}^C > \tilde{W}^{ET} > \tilde{W}^{RPM}$	$\tilde{W}^C > \tilde{W}^{ET} > \tilde{W}^{RPM}$	$\tilde{W}^C \gtrless \tilde{W}^{ET}$ (*RPM* dominated by both)

tailers' shocks is irrelevant and previous formulas for the *ET* and *RPM* cases are still valid. The main difference between idiosyncratic and market uncertainties is that, under idiosyncratic uncertainty and competition, cost differentiation leads to some double marginalization and retailers profit variability. Although competition still keeps a lid on retail prices (at least under our assumption that cost differentials are not too big), it no longer functions as a perfect tournament.

V. When do Private and Social Objectives Conflict?

Table 2 summarizes our results about the private and social desirabilities of vertical restraints for undifferentiated retailers.[10]

The main policy issue in our model comes from the fact that consumers are left out of the contract between the manufacturer and the retailers. While the contract is efficient from the point of view of the parties that sign it, the externality on the consumers may call for public intervention.[11]

The analysis shows that private and social objectives may well conflict. For instance, when the retailers are risk neutral, the manufacturer may impose a vertical restraint (*ET*) and the social planner (or antitrust authority) may want to restore competition between the retailers.

There seems to be a general bias in our model for the *consumers to prefer competition to ET and RPM.* This may seem trivial. As competition reduces the retailers' profit margins, one might think that it therefore lowers consumer price. However, this reasoning is false, since it ignores the difference in wholesale price between regimes. Indeed, in the absence of uncertainty, the increase in wholesale price exactly offsets the decrease in the profit margin (Proposition 1) so that the consumers are unaffected by vertical restraints. The reasons why the consumers in general prefer competition are in our model connected to uncertainty.

The consumers' preference for competition has two origins. Remember that the consumers' surplus ($S = E\{(d-q)^2/2\} = ((d^e - q^e)^2 + \mathrm{var}(d-q))/2$) decreases with the average price and increases with the variance of consumption (since it is convex). For undifferentiated products and market uncertainty, competition scores better on the two accounts:

---

[10] For tie-breaking, see Sections II and III.

[11] Mathewson-Winter (1983a,b,c) show in a wide variety of circumstances that, when there is no uncertainty, vertical restraints increase *aggregate* welfare. Gallini-Winter exhibit a case in which aggregate welfare can be higher without *any* restraint (in particular, the manufacturer is prevented from charging a franchise fee) than with *all* restraints. However, this result does not hold if franchise fees are allowed, as is usually the case under current U.S. law. Indeed, a franchise fee in their model is a sufficient instrument, so that the manufacturer does not lose anything by not using *ET* or *RPM* even if he can do so.

1) The expected price is lower under competition.

2) The competitive price is more sensitive to cost disturbances and less sensitive (or at least not more sensitive) to demand disturbances than the *ET* and *RPM* prices.[12] So in both cases, the variance of consumption is higher under competition.

Let us provide some intuition about why this is so.

1′) Let us see why the expected price is lower under competition. Again, it is useful to start from the no uncertainty case, in which the consumer price is the same for the three regimes. Let us introduce market uncertainty and retailers' risk aversion. The manufacturer then may want to change the wholesale price to reduce the retailers' risk (*remember that the ex post retailing behavior is independent of the ex ante uncertainty and risk aversion, so that we can identify changes in the wholesale price with changes in the consumer price*). Under competition, the retailers' profit is independent of the wholesale price and therefore a wholesale price adjustment does not help reducing risk. To the contrary, under *ET* or *RPM*, the manufacturer can in general decrease the retailers' risk by increasing the wholesale price (and decreasing correspondingly the franchise fee): a lower profit margin means a lower variance of profits, as can easily be checked. So uncertainty calls for an increase in the wholesale price only when the contract specifies an anticompetitive restraint. Again, lower expected consumer price under competition can be attributed to the good insurance property of the competitive system. Under retailers' risk neutrality, however, competition, *ET* and *RPM* lead to the same expected price, as could be expected from the previous reasoning.

2′) Let us now see why consumption has a higher variance under competition. Consider first the case of demand uncertainty. The competitive price is entirely determined by cost conditions (from our constant returns to scale technology); therefore, it does not adjust to demand shocks, nor does it under *RPM*. But under *ET*, the consumer price partially adjusts to demand conditions. So consumption varies most under competition and *RPM*. Second, consider the case of cost disturbances. The competitive price adjusts 100 percent with the retail cost, while it adjusts partially under *ET* and not at all under *RPM*. Again, the variance of consumption is highest under competition.

These two effects—lower expected prices and higher variance of consumption—create a bias in the policy intervention towards favoring competition. Indeed, the social planner may want to prohibit a privately desirable restraint.

The case of idiosyncratic uncertainty is a bit less clear-cut. The idiosyncracy reintroduces monopoly profits and risk under competition. Thus the general analysis becomes more complex. One can nevertheless find reasonable conditions on the distribution of uncertainty for extreme risk aversion, such that, for example, the consumers prefer competition because the competitive price both is lower on average and fluctuates more than the price under a vertical restraint.

A word of caution should be added here. Our conclusions hold only if there is effective competition between the retailers in the absence of restraints. If the retailers can collude, the competitive solution closely resembles the *ET* one; and preventing the manufacturer from imposing anticompetitive restraints (assuming he still wants to do so) loses much of its interest.

VI. General Retailer Risk Aversion and Demand Function

We now generalize a number of our results to more general objective functions for the retailers, and some of these results to arbitrary demand functions as well. Each retailer has a von Neumann-Morgenstern, differentiable, increasing, and concave utility function U. His reservation utility is $U(0)$. The "general" demand function is denoted by $D(q,d)$. It is strictly decreasing and differentiable with respect to the retail price.

---

[12] Because retailers are not differentiated, the consumer price reacts less to demand disturbances under *RPM*.

Throughout this section, we consider only *undifferentiated retailers* and *market uncertainty* (as noted earlier, our points about the effects of competition are made most strongly in this pure Bertrand case).

Let us first study the *private desirability* of vertical restraints. Sections II and III showed that, for linear demand, the manufacturer prefers competition to *ET* under extreme risk aversion and *ET* to competition under risk neutrality. Proposition 4 below shows that, for a general demand function, when the retailers' Arrow-Pratt index of absolute risk aversion increases, the manufacturer favors competition over *ET* more and more.

PROPOSITION 4: *Let* U_0 *and* U_1 *denote two utility functions for the retailers and let* $\{\Pi_0^C, \Pi_0^{ET}, \Pi_1^C, \Pi_1^{ET}\}$ *denote the resulting profits for the manufacturer under competition and ET, for a general demand function. Then, under demand and/or cost uncertainty, if* $(-U_1''/U_1') > (-U_0''/U_0')$,

$$\text{(a) } \Pi_1^C = \Pi_0^C \qquad \text{(b) } \Pi_1^{ET} < \Pi_0^{ET}.$$

PROOF:

(a) A simple, but important step in the proof of this proposition (like in subsequent proofs) consists in noticing that, under competition, the retailers are perfectly insured, so that the retailers' degree of risk aversion is irrelevant. This holds for any demand function (in particular, for linear demand, formulas (1) through (4) hold regardless of U).

(b) Let us now show that $\Pi_1^{ET} < \Pi_0^{ET}$. This later property actually is a simple consequence of the Arrow-Pratt theorem as we demonstrate shortly.

Consider exclusive territories and let p_1 and A_1 denote the optimal wholesale price and franchise fee when the retailers' utility function is U_1. Let $V_r(p, d, \gamma)$ and $V_m(p, d, \gamma)$ denote a retailer's and the manufacturer's *variable* profits (V_m is a profit per retailer. For linear demand: $V_r(p, d, \gamma) = \frac{1}{8}(d - \gamma - p)^2$ and $V_m(p, d, \gamma) = \frac{1}{4}(d - \gamma - p)(p - c)$). The manufacturer's profit per retailer for an *ET* contract $\{p, A\}$ is $\{A + EV_m(p, d, \gamma)\}$, where expectations are taken with respect to the demand and cost parameters. Also, in order for a retailer to accept the contract:

$$EU(-A + V_r(p, d, \gamma)) \geq U(0). \tag{26}$$

For utility function U_1, the optimal contract $\{p_1, A_1\}$ satisfies (26) with equality. Thus, zero is the certainty equivalent of the gamble $\{-A_1 + V_r(p_1, d, \gamma)\}$. Suppose now that the retailers' utility function is U_0. Then, from the Arrow-Pratt theorem, the certainty equivalent of the gamble $\{-A_1 + V_r(p_1, d, \gamma)\}$ strictly exceeds zero. Thus, the contract $\{p_1, A_1\}$ is accepted for utility U_0; and leaves a rent to the retailer. Hence, for ε small, the contract $\{p_1, A_1 + \varepsilon\}$ is also accepted. This proves that $\Pi_0^{ET} > \Pi_1^{ET}$, since the retailers' *ex post* decisions depend only on the wholesale price and not on the degree of risk aversion.

We next show that for any degree of risk aversion, the manufacturer prefers competition to *RPM* (as long as retailers are not differentiated). The proof of Proposition 5 requires the demand function to be (weakly) convex (as is the case for instance for linear and exponential demands).

PROPOSITION 5:

(a) *For demand uncertainty*, $\hat{\Pi}^C = \hat{\Pi}^{RPM}$.

(b) *For cost uncertainty, if* U *is strictly concave, and* D *is convex in* q, $\tilde{\Pi}^C > \tilde{\Pi}^{RPM}$.

PROOF:

(a) Under demand uncertainty and competition, fixing the wholesale price p amounts to fixing the retail price $q = p + \gamma$. Furthermore, the best wholesale price given a fixed *RPM* price q is $p = q - \gamma$, because it gives perfect insurance to the retailers, like in the competitive case. So, the two arrangements are equivalent.

(b) Consider cost uncertainty. Let $\{A^{RPM}, q^{RPM}, p^{RPM}\}$ denote the optimal contract under *RPM* for an arbitrary strictly concave function U. Suppose that the manufacturer offers a competitive arrangement, with $\{p^C \equiv q^{RPM} - \gamma^e, A^C \equiv 0\}$. The retailers, who make a zero profit in all states of nature, accept the competitive contract.

Under *RPM*, a retailer has variable profit $\frac{1}{2}D(q^{RPM})(q^{RPM}-p^{RPM}-\gamma)$, which must satisfy

$$E_\gamma U\big(-A^{RPM}+\tfrac{1}{2}D(q^{RPM}) \times (q^{RPM}-p^{RPM}-\gamma)\big)=U(0),$$

which, together with the strict concavity of U, implies

$$(27)\quad A^{RPM}<\tfrac{1}{2}D(q^{RPM}) \times (q^{RPM}-p^{RPM}-\gamma^e),$$

which, in turn, implies that the manufacturer's profit satisfies

$$(28)\quad 2A^{RPM}+D(q^{RPM})(p^{RPM}-c) < D(q^{RPM})(q^{RPM}-\gamma^e-c).$$

Under the proposed competitive scheme, the manufacturer's expected profit is

$$E_\gamma\big(D(q^{RPM}-\gamma^e+\gamma)(q^{RPM}-\gamma^e-c)\big).$$

But, if the demand function is convex,

$$(29)\quad E_\gamma D(q^{RPM}+\gamma-\gamma^e)>D(q^{RPM}).$$

Thus, the manufacturer does better under the competitive scheme.

Proposition 5 shows that, for general uncertainty and risk aversion and convex demand, the manufacturer chooses between competition and *ET* (assuming all contracts are available). Proposition 4 and Sections II and III show that, for general uncertainty and risk aversion and linear demand, the manufacturer prefers *ET* if retailers are not very risk averse and competition if they are risk averse.

Finally, let us consider the issue of *social desirability*. We have already noted a number of cases in which private and social objectives conflict. Let us now show that, for general risk aversion and linear demand, consumers prefer competition to the choice of a competition-reducing restraint:

PROPOSITION 6: *Assume demand is linear:* $D(q,d)=d-q$; *then for any degree of retailers' risk aversion, expected consumers' surplus is higher under competition than under ET or RPM.*

PROOF:

As discussed in Section V, the expected consumers' surplus decreases with the expected consumer price, and increases with the variance of consumption. We show that competition scores better on both counts. We already noticed in Section V that the competitive retail price fully responds to the retail cost and does not respond to the demand parameter. Thus, consumption reacts at least as much to cost and demand shocks under competition as under *ET* or *RPM*.

Therefore, we need only show that the average consumer price is lower under competition. To this purpose, let us start with the case of risk-neutral retailers. For a linear demand, (3), (20), and (23) show that the expected consumer price is the same under the three arrangements. Now introduce retailers' risk aversion. As we mentioned, the consumer price is unaffected under competition.

Consider next *ET*. To show that the expected consumer price is higher under risk aversion, it suffices to show that the wholesale price is higher than under risk neutrality: $p^{ET}\geq c$. Suppose that $p^{ET}<c$. Let us show that an exclusive territory contract with a wholesale price equal to c would make the manufacturer better off. To this purpose, let $A(p)$ be the maximum franchise fee that the manufacturer can impose together with wholesale price p:

$$(30)\quad E_g U(-A(p)+V_r(p,g))=U(0),$$

where $g\equiv d-\gamma$ and $V_r(p,g)=\frac{1}{8}(g-p)^2$ is each retailer's variable profit. Differentiating (30) gives

$$(31)\quad [-A'(p)] = \frac{E(U'(-(\partial V_r/\partial p)))}{EU'}.$$

(where $A' < 0$ and $\partial V_r/\partial p < 0$). Using the fact that the covariance with respect to g of U' and $(-(\partial V_r/\partial p))$ is negative, (31) implies

$$(32)\quad [-A'(p)] \leq E[-(\partial V_r/\partial p)].$$

Thus, for $p^{ET} < c$,

$$(33)\quad A(p^{ET}) - A(c) \leq \int_{p^{ET}}^{c} (E(-(\partial V_r/\partial p)))\, dp.$$

We also know that

$$(34)\quad E(V_m(c,g) + V_r(c,g)) > E(V_m(p,g) + V_r(p,g)),$$

for any p differing from c ((34) is simply implied by the fact that residual claimancy for the retailer maximizes aggregate profit in all states of nature). Equations (33) and (34) then imply that

$$\begin{aligned}(35)\quad & [A(c) + E(V_m(c,g))] - [A(p^{ET}) + E(V_m(p^{ET},g))] \\ & \geq \int_{p^{ET}}^{c} (E(-(\partial V_r/\partial p)))\, dp - [E(V_r(p^{ET},g)) - E(V_r(c,g))] \\ & = \int_{p^{ET}}^{c} (E(-(\partial V_r/\partial p)))\, dp - E\left(\int_{p^{ET}}^{c} (-(\partial V_r/\partial p))\, dp\right) = 0.\end{aligned}$$

Thus, we conclude that a wholesale price lower than c under ET is not optimal.

Lastly, let us consider *RPM*. Under demand uncertainty we know that *RPM* is equivalent to competition. Thus, let us consider cost uncertainty. Again, if the retailer is risk neutral, the consumer price, q^* say, is such that *aggregate* expected profit is maximized. q^* is also the expected price under competition. Let us show that, under risk aversion, $q^{RPM} \geq q^*$. To this purpose, let us consider a family of *RPM* contracts such that $(q-p)$ is constant. This will enable us to use a reasoning similar to the one for *ET*. Let us for this family of contracts introduce the following notation: $V_r(q,\gamma)$ and $V_m(q)$ denote each retailer's profit and the manufacturer's profit per retailer, when p is adjusted so as to keep $(q-p)$ constant (for instance, in the linear demand case, $V_r(q,\gamma) = \frac{1}{2}(d-q)(k-\gamma)$ and $V_m(q) = \frac{1}{2}(d-q)\times(q-k-c)$, where $k \equiv q-p$ is a constant).

Let $A(q)$ be defined by

$$(36)\quad EU(-A(q) + V_r(q,\gamma)) = U(0),$$

where expectations are taken with respect to γ. Differentiating (36) and using the negative covariance between U' and $(-(\partial V_r/\partial q))$ gives:

$$(37)\quad [-A'(q)] = \frac{E(U'(-(\partial V_r/\partial q)))}{EU'} \leq E(-(\partial V_r/\partial q)),$$

which implies

$$(38)\quad A(q^{RPM}) - A(q^*) \leq \int_{q^{RPM}}^{q^*} E(-(\partial V_r/\partial q))\, dq$$

if $q^{RPM} < q^*$.

Next, we observe that q^* maximizes $E[V_r(q,\gamma) + V_m(q)]$. So for all q,

$$(39)\quad E[V_r(q,\gamma) + V_m(q)] < E[V_r(q^*,\gamma) + V_m(q^*)].$$

Using (38) and (39), we obtain, for $q^{RPM} < q^*$,

$$\begin{aligned}(40)\quad & [A(q^{RPM}) + V_m(q^{RPM})] \\ & - [A(q^*) + V_m(q^*)] \\ & < \int_{q^{RPM}}^{q^*} E(-(\partial V_r/\partial q))\,dq \\ & - E\left(\int_{q^{RPM}}^{q^*} (-(\partial V_r/\partial q))\,dq\right) = 0.\end{aligned}$$

This shows that a *RPM* contract with $q^{RPM} < q^*$ cannot be optimal for the manufacturer. The latter could raise his profit by offering an alternative *RPM* contract with resale price q^*. We thus conclude that the expected consumer price under *RPM* exceeds q^*, that is, exceeds the expected consumer price under competition.

We now conclude by showing that, under linear demand, competition is always socially optimal:

PROPOSITION 7: *Under linear demand, for any degree of retailer risk aversion, aggregate welfare under competition exceeds those under ET and RPM.*

PROOF:

As before, retailers are put at their reservation utilities, so that their welfare needs not be taken into account for welfare comparisons. Thus, we want to show that, under linear demand,

$$(41)\quad \Pi^C + S^C > \max\{\Pi^{ET} + S^{ET}, \Pi^{RPM} + S^{RPM}\}.$$

That welfare is higher under competition than under *RPM* is a direct consequence of Propositions 5 and 6, which show that both the manufacturer and the consumers prefer competition to *RPM*.

Let us now show that welfare is higher under competition than under *ET*. First notice that

$$(42)\quad S^C - S^{ET} = \left[\tfrac{1}{8}(g^e - c)^2 + \tfrac{1}{2}\sigma_g^2\right] - \left[\tfrac{1}{8}(g^e - p^{ET})^2 + \tfrac{1}{8}\sigma_g^2\right],$$

where $g = d - \gamma$. From the proof of Proposition 6, $p^{ET} \geq c$. Equation (42) then implies that

$$(43)\quad S^C - S^{ET} \geq \tfrac{3}{8}\sigma_g^2.$$

To complete the proof, notice that, from Proposition 4, $(\Pi^{ET} - \Pi^C)$ is highest when the retailers are risk neutral; and that for risk-neutral retailers, from (21) and (4), we have: $\Pi^{ET} - \Pi^C = \frac{1}{4}\sigma_g^2$.

Proposition 7 implies that for market uncertainty, undifferentiated retailers and linear demand, anticompetitive restraints should always be prohibited.

VII. Conclusion

The following conclusions emerge from our analysis:

1) Under incomplete information about the retailers' environment, the manufacturer may not want to impose vertical restraints, even if he (or she) possesses the information required to do so. The efficiency in using decentralized information and the insurance properties of vertical restraints have to be compared with those of the competitive policy. On the whole, the competitive policy scores well for market uncertainty and extreme risk aversion, because its "tournament" aspect insures the retailers. It does not do as well when the retailers are not very risk averse.

When the consumer price is not observable, or at least not verifiable by a court, the manufacturer's choice between *ET* and competition depends on the retailers' risk aversion. As noted in Section I, Part B, the choice of consumer price by a retailer is but a form of moral hazard. Alternatively, one could assume that the manufacturer is able to monitor the consumer price, but not some selling effort levels exerted by the retailers. The manufacturer then faces the same tradeoff between an efficient, because unconstrained by product competition, use of decentralized information (*ET*), and and efficient provision of insurance, that is, a better capture of the retailers' informational rent (product competition). In this sense, our paper studies product competition as an incentive device. It differs from the contributions by Oliver Hart (1983) and David Scharfstein (1985) in that competition occurs between agents working for the same principal, and thus is a choice variable for the latter.

2) A major contribution of the earlier literature on vertical restraints is to have shown that *RPM* and *ET* can be substitutes (R. Bork, 1965–66). Here we shade this view somewhat by finding that the two policies behave fairly differently under uncertainty. The two standpoints actually are not inconsistent as they address different problems. The earlier literature emphasized that *RPM* and *ET* both insulate retailers from intrabrand competition. Thus, they both can be used to solve the free-rider problems in local sale efforts and postsale services. This paper ignores the free-rider issue, and focuses on the delegation problem under incomplete information. *ET* clearly uses decentralized information more efficiently than *RPM*. But *RPM* has better insurance properties in the case of demand uncertainty. The search for the best control policy for the manufacturer (when restraints are feasible and effective) must rest on a fine analysis of the type of uncertainty (cost/demand, market/idiosyncratic) and of the retailers' degree of risk aversion.

3) Another major contribution of the earlier literature on vertical restraints is to have shown that per se illegality of such restraints has no economic foundations (for example, Spengler; Telser; Richard Posner, 1976, 1977; Oliver Williamson, 1975; Mathewson-Winter, 1983a; Blair-Kaserman). The legal status of vertical restraints is fairly involved. Roughly, in the United States, competition (franchise fees) is presumptively legal; *ET*, after being illegal per se, is now subject to the rule of reason; and *RPM* suffers from per se illegality except in states which have passed a fair trade law. Recently, Posner (1981) has suggested that vertical restraints ought to be legal per se (unless they result from a dealers' cartel, which is not the case in this paper). Our paper does not support this claim. The planner may well favor competition over the manufacturer's choice of a vertical restraint. Section V and Propositions 6 and 7 emphasize the social planner's bias in favor of competition. So at the current stage of research, the rule of reason seems safer. As Posner (1977) notes, however, the rule of reason places a heavy burden of analysis on the antitrust judge. Analysis such as the one presented in this paper may in the long run provide some guidelines for the application of this rule.

Much work remains to be done. First, this paper focuses on the interaction between the use of decentralized information, double marginalization and insurance. We did not consider other important vertical issues such as the provision of selling effort (see above item 1, though). Second, another yet missing piece of the theory ought to explain why the antitrust authorities commonly use only elementary controls over vertical structures such as the prohibition of vertical restraints. The antitrust authorities' limited information about the environment will be a key to answer this query. Third, we have assumed that there is a competitive supply of retailers. This may be reasonable as long as retailers invest little in their relationship with the manufacturer and if they do not possess attributes, such as location and facilities, that make them special to the manufacturer. Otherwise retailers have some bargaining power (see Williamson). The outcome may move towards a dealers' cartel. The study of retailers' bargaining power can provide a link between the theories of vertical and horizontal restraints. Lastly, we would want to consider the link between competition at the upstream level and vertical restraints. All these questions must await further treatment.

APPENDIX

Idiosyncratic Uncertainty

Here we give the results only for a uniform cost distribution on $[\underline{\gamma}, \bar{\gamma}]$. The same results can be shown to hold for general exponential distributions.

Let us first assume that the retailers are *infinitely risk averse*. It can easily be shown that, as in the market uncertainty case, *RPM* is inferior to competition and *ET* from both the manufacturer's and the social planner's points of view. So we compare competition and *ET*.

1. *Competition*. The franchise fee must equal zero as each retailer makes zero profit for some states of demand.

Note that, if the cumulative distribution function of each retailer's cost γ_i is $F(\cdot)$, the cumulative distribution function of $\gamma^{\circ} = \sup\{\gamma_1, \gamma_2\}$ is given by $F^{\circ}(\cdot) = F^2(\cdot)$. Straightforward computations then lead to

$$\tilde{A}^C = 0 \tag{A1}$$

$$\tilde{p}^C = \tfrac{1}{2}(d + c - \gamma^{oe}) \tag{A2}$$

(where $\gamma^{oe} \equiv E\gamma^{\circ} = \frac{2}{3}\bar{\gamma} + \frac{1}{3}\underline{\gamma} > \gamma^e$, for the uniform distribution),

$$\tilde{q}^C = \tfrac{1}{2}(d + c - \gamma^{oe}) + \gamma^{\circ} \tag{A3}$$

$$\tilde{\Pi}^C = \tfrac{1}{4}(d - c - \gamma^{oe})^2 \tag{A4}$$

$$\tilde{W}^C = \tfrac{3}{8}(d - c - \gamma^{oe})^2 + \tfrac{1}{2}\sigma_\gamma^{o2} \tag{A5}$$

(where $\sigma_\gamma^{o2} \equiv E(\gamma^{\circ} - \gamma^{oe})^2 = \Delta^2/18 < \sigma_\gamma^2 = \Delta^2/12$, for the uniform distribution, and $\Delta \equiv \bar{\gamma} - \underline{\gamma}$).

2. *Exclusive Territories*. Under *ET*, the outcome is the same as in the case of market uncertainty for each realization of the cost parameter: the degree of correlation is irrelevant as long as territories are exclusive. So $\{\tilde{A}^{ET}, \tilde{p}^{ET}, \tilde{q}^{ET}, \tilde{\Pi}^{ET}, \tilde{W}^{ET}\}$ are given by equations (6) through (10).

It is possible to show that, under our assumption that $\bar{\gamma}$ is not too high relative to $\underline{\gamma}$, (9), (10), (A4), and (A5) imply that $\tilde{\Pi}^C > \tilde{\Pi}^{ET}$ and $\tilde{W}^C > \tilde{W}^{ET}$. The consumers prefer competition as from (8) and (A3),

(*i*) $E\tilde{q}^{ET} - E\tilde{q}^C = \frac{1}{2}(\bar{\gamma} - \gamma^{oe}) > 0$;

(*ii*) $\sigma_\gamma^{o2} > \frac{1}{4}\sigma_\gamma^2$ (the consumer price under competition (which fluctuates like γ°) has a higher variance than the consumer price under *ET* (which fluctuates like $\gamma/2$)). See Section V for a broader discussion of consumers' welfare.[13]

If retailers are *risk neutral*, these results are no longer valid. Actually in this case, both the manufacturer and the social planner prefer *ET* to competition and competition to *RPM*. To check this point, it suffices to note that in the case of *ET*, the nature (idiosyncratic or not) of uncertainty is not relevant when deriving the optimal tariff policy. Since *ET* allows the manufacturer to realize the vertically integrated profit under risk neutrality, $\tilde{\Pi}^{ET} > \tilde{\Pi}^C$. So the manufacturer imposes a vertical restraint. It is also possible to show that, for an exponential distribution, $\tilde{W}^{ET} > \tilde{W}^C$. (The expected consumer price is lower under *ET*, but the variance of the consumer price is higher under competition.)

Thus, for idiosyncratic uncertainty, no conflict between private and social objectives arises if we restrict ourselves to exponential uncertainty and one of the two polar cases of risk aversion. But the analysis clearly suggests that conflicts will arise if we look at more general specifications.

---

[13] While (*i*) holds for any distribution, (*ii*) may not hold for distributions that are not well-behaved (for instance, (*ii*) is not satisfied by a two-point distribution with most of the weight on the high value).

REFERENCES

Blair, Roger and Kaserman, David, *Law and Economics of Vertical Integration and Control*, New York: Academic Press, 1983.

Bork, Robert, "The Rule of Reason and the Per se Concept; Price Fixing and Market Division: I and II," *Yale Law Journal*, Winter 1965, *74*, 775–847; January 1966, *75*, 373–475.

Dixit, Avinash, "Vertical Integration in a Monopolistically Competitive Industry," *International Journal of Industrial Organization*, 1983, *1*, 63–78.

Gallini, Nancy and Winter, Ralph, "On Vertical Control in Monopolistic Competition," *International Journal of Industrial Organization*, September 1983, *1*, 275–86.

Hart, Oliver, "The Market Mechnism as an Incentive Scheme," *Bell Journal of Economics*, Autumn 1983, *14*, 366–82.

Mathewson, G. F. and Winter, Ralph, (1983a) "The Economics of Vertical Restraints in Distribution," in I.E.A. Conference Volume: *New Developments in the Analysis of Market Structures*, Cambridge: MIT Press, 1983.

______ **and** ______, (1983b) "The Incentives for Resale Price Maintenance under Imperfect Information," *Economic Inquiry*, July 1983, *21*, 337–48.

______ **and** ______, (1983c) "Vertical Integration by Contractual Restraints in Spatial Markets," *Journal of Business*, October 1983, *56*, 495–517.

______ **and** ______, "An Economic Theory of Vertical Restraints," *Rand Journal of Economics*, Spring 1984, *15*, 27–38.

Mookherjee, Dilip, "Optimal Incentive Schemes with Many Agents," *Review of Economic Studies*, July 1984, *21*, 433–46.

Posner, Richard, *Antitrust Law: An Economic Perspective*, Chicago: University of Chicago Press, 1976.

______, "The Rule of Reason and the Economic Approach: Reflections on the Sylvania Decision," *University of Chicago Law Review*, Fall 1977, *45*, 1–20.

______, "The Next Step in the Antitrust Treatment of Restricted Distribution: Per se Legality," *University of Chicago Law Review*, Winter 1981, *48*, 6–26.

Rey, Patrick, and Tirole, Jean, "Vertical Restraints from a Principal-Agent Viewpoint," in L. Pelligrini and S. Reddy, eds., *Marketing Channels: Relationships and Performance*, Lexington: Lexington Books, 1986.

Scharfstein, David, "Product Market Competition and Managerial Slack," mimeo., MIT, 1985.

Scherer, F. M. , *Industrial Market Structure and Economic Performance*, 2nd ed., Boston: Houghton Mifflin, 1980.

Spengler, Joseph, "Vertical Integration and Anti-trust Policy," *Journal of Political Economy*, August 1950, *58*, 347–52.

Telser, Lester, "Why should Manufacturers Want Fair Trade?," *Journal of Law and Economics*, October 1960, *3*, 86–105.

Williamson, Oliver, *Markets and Hierarchies*, New York: Free Press, 1975.

Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation

By JOSEPH FARRELL AND GARTH SALONER*

A good is often more valuable to any user, the more others use compatible goods. We show that this effect may inhibit innovation. If an installed base exists and transition to a new standard must be gradual, early adopters bear a disproportionate share of transient incompatibility costs. This can cause "excess inertia." The installed base, however, is "stranded" if the new standard is adopted: this may create "excess momentum." These dynamic effects have strategic implications.

When compatibility is important, an installed base of durable goods or training may affect the likelihood and desirability of innovation. In this paper we analyze the private and social incentives for the adoption of a new technology that is incompatible with the installed base.

The benefits from compatibility create demand-side economies of scale: there are benefits to doing what others do. These benefits make standardization a central issue in many important industries. There are three main sources of these benefits. The first is *interchangeability of complementary products* such as computer software, VCR tapes, phonographic records, cassette tapes, and camera lenses. The second is *ease of communication* (between people or between people and machines). The most important example is that of telecommunications networks: the value to each telephone subscriber depends on the number of other people on the network. Other examples include standardized typewriter keyboards or machinery (this makes learned skills more widely usable), weights and measures, and, of course, language itself. The third is *cost savings*: standardization, especially interchangeability of parts, facilitates mass production.[1]

In the presence of compatibility benefits, a user who switches to a new, superior technology cannot obtain its full benefit unless other current users also switch and new users adopt the new technology. This creates the possibility of "excess inertia": a socially excessive reluctance to switch to a superior new standard when important network externalities are present in the current one.

In an earlier paper (1985), we examined this problem in a model in which all users have the opportunity to adopt the new technology at essentially the same time, that is, payoffs depend only on *who* switches, not on *when* they switch. We found that symmetric excess inertia (a Pareto-superior new technology not being adopted) could not occur with complete information,[2] although it could occur with incomplete information.[3]

---

*GTE Laboratories, 40 Sylvan Road, Waltham, MA 02254 (visiting UC-Berkeley, 1986–87), and Department of Economics, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02139 (visiting Graduate School of Business, Stanford University and Hoover Institution, 1986–87), respectively. Saloner thanks the National Science Foundation, grant no. IST-8510162, for financial support. We both thank the referees for helpful comments.

[1] This categorization is not exhaustive. For example, servicing of automobiles is cheaper and easier to obtain for common models because they are more familiar to mechanics and because spare parts are more widely available. (See our 1985 paper and Michael Katz and Carl Shapiro, 1985, for other examples.)

[2] The intuition for this result can be obtained by considering a sequence of N decision makers contemplating switching, and using a backwards induction argument. If everyone else has switched, the Nth decision maker will choose to switch. But then the $(N-1)$st decision maker can be certain that it will be followed in a switch by the Nth decision maker if *its* predecessors have all switched. Thus it too will find it optimal to switch, and so on. (See our 1985 paper, Sec. 2.)

[3] With incomplete information about the benefit of the network externalities to others, no firm can be sure

In practice, however, not all users have the same opportunities at a given time: when innovation is unexpected, new users have an option that was unavailable to previous users, and, moreover, those (the installed base) who had previously adopted the old technology may be at least somewhat *committed*. As a result, if the technology is adopted, it will take time for its network to grow.

These delays in building a network on the new standard can create inefficiencies. An early adopter of the new technology creates at least transient incompatibility. While this is a social cost of the innovation, the private costs and benefits may not accurately reflect the social ones. In our earlier paper (1985), any permanent incompatibility was important, but, since we ignored interim payoffs, any *transient* incompatibility did not matter. The two models we analyze here, on the contrary, emphasize that effect.

In the first model, we assume that the network on the new technology is built up of new users, and that old users who switch do not significantly contribute to it. In our second model, we assume by contrast that new users are negligible, and that the new network is built up through old users switching. The delays in building a substantial new network are caused in the first case by the time it takes for enough new users to arrive, and in the second by the time it takes for old users to switch. Before going to the formal treatment, we informally discuss the two models.

A Model with New Users. First, we consider the case in which the new network must be built up through adoption by new users. This is a reasonable assumption if old users switch slowly, compared to the arrival rate of new users. We suppose that there is a continual stream of (infinitesimal) new users. We examine what determines whether a new (and unexpected) technology gets adopted and emerges as the new standard.

The following examples may illustrate what we have in mind: there was a large installed base of "Standard 8mm" movie cameras and projectors when "Super 8" was introduced; the "QWERTY" keyboard was ubiquitous by the time the Dvorak keyboard became available (see Paul David, 1985); and the motor car was invented at a time when there was a substantial installed base of horse-drawn carriages and streetcars.

As we show in Section I, the presence of the installed base causes a disparity between the social incentives for the adoption of the new technology and the private incentives facing individual decision makers. There are two externalities. *First*, adoption of the new technology affects the users of the old technology. Their network ceases to grow, and may even shrink as some current users abandon their old equipment for newer equipment that uses the new technology. For instance, when Super 8 was introduced, recent buyers of Standard 8 equipment found that Standard 8 film became harder to obtain, and delays in processing grew. They suffered a loss, therefore, from the introduction of Super 8—especially relative to the benefits they had expected. This loss was not taken into account by the sellers or the buyers of the new Super 8 technology. *Second*, an early adopter of the new technology enhances its appeal to later users (and reduces the appeal of the old). In general, he or she does not appropriately take this into account.

In our model we consider the game between users who arrive at different times. Each decides which technology to adopt, given what others have done. The equilibrium outcome depends on the size of the installed base when the new technology is introduced, how quickly the network benefits of the new technology are realized, and the relative superiority of the new technology. The parameter space representing these attributes can be divided into three regions: In the first, the unique perfect Nash equilibrium is that the new technology is adopted; in the second, it is that it is not adopted; and in the third, both these outcomes are equilibria.[4]

---

that it would be followed in a switch to the new technology. This uncertainty can lead all the firms to remain with the status quo even when in fact they all favor switching, because they are unwilling to risk switching without being followed.

[4] The multiple equilibria arise from the fact that the outcome may depend on new adopters' expectations about what other adopters will do. If each user expects everyone (no one) else to adopt he will usually also

Because there are unique equilibria in some regions, we can illustrate unambiguously how the two externalities discussed above can lead to inefficient technology choice. Suppose, for example, that the new technology is far superior to the old in the long run (i.e., once the network is large), but that it would take a long time for the network to become established. Early adopters of the new technology would bear a disproportionate share of the transient incompatibility costs. They may be unwilling to do this. In that case, the unique equilibrium is that the new technology is not adopted, despite the long-run benefits. This is more likely to occur if the installed base is large, or if the new technology is unattractive when it has only a few users. Thus the installed base may cause excess inertia. The Dvorak keyboard's failure to displace the less efficient QWERTY keyboard may be an example of this.[5] Once the first potential adopters decide not to adopt the new technology, of course, the excess inertia only becomes worse.

Excess momentum—the inefficient adoption of a new technology—can also arise. Suppose that the new technology offers the first potential adopter an advantage over the current technology. He (or she) may be willing to adopt the new technology even though it will be a long time before the network is established. Once he adopts it, the new technology becomes even more attractive for later users, and the unique equilibrium may be that it is adopted. Excess momentum may result, because of the "stranding" (or "orphaning") externality, which hurts the users comprising the installed base.

Because the size of the installed base may critically affect adoption, it may constitute a *barrier to entry*. Not only does the presence of a barrier affect efficiency, but firms will have an incentive to take actions to buttress it. One such action that has been alleged in antitrust litigation is the announcement of future availability of a new product.[6] Defendants have been charged with making a "premature announcement" or a "predatory preannouncement" in order to discourage existing customers from switching to another supplier and to encourage those intending to buy soon to wait, and thus not become part of the "installed base." No formal models have been developed to deal with this question. However, several authors have claimed that preannouncements cannot be anticompetitive. For example:

> In general, there is no reason to inhibit the time when a firm announces or brings products to the marketplace. Customers will be the final arbiter of the product's quality and the firm's reputation.... Advance announcements of truthful information about products cannot be anticompetitive. Indeed, such announcement is procompetitive; competition thrives when information is good.
>
> [Franklin Fisher, John McGowan, and Joen Greenwood, p. 289][7]

However, when there are significant network externalities, the timing of the announcement of a new incompatible product can critically determine whether the new product supersedes the existing technology. In that case, because of the externalities arising from the installed base, a preannouncement can sometimes secure the success of a new technology that is socially not worth adopting, and that would not have been adopted absent the preannouncement.

The intuition for this result is the following: With a preannouncement, two effects favor the new technology. First, if some users decide to wait for it, the network benefits

---

adopt (not adopt). Imposing rationality on these expectations sometimes leads to a unique equilibrium, but not always. Where multiple equilibria exist, which equilibrium prevails may depend on which is more "focal" (see Thomas Schelling, 1960, for a discussion of this). The incentive to make an outcome more focal than another may lead to large investments in advertising and to "introductory" pricing.

[5] See David for the history of the typewriter keyboard.

[6] See Franklin Fisher, John McGowan and Joen Greenwood (1983) for a discussion of "predatory preannouncements" in the IBM case.

[7] Similarly, Robin Landis and Ronald Rolfe state: "...the welfare of consumers can only be increased by having additional correct information that is relevant to their purchasing decision" (1985, p. 140).

when the new technology is introduced (and adopted by those users) will be larger than otherwise. Second, the installed base on the old technology will be reduced by the number of users who wait. In some cases, the unique equilibrium without a preannouncement is that the new technology is not adopted, while with a preannouncement it is adopted. Of course, the potential users who decide to wait are indeed well-informed "arbiters of the product's quality" and their welfare is increased, but they are not the only ones who matter. Their adoption of the new technology affects both the users in the installed base and later adopters who might have preferred the old technology to the new. Thus, the preannouncement may reduce welfare.

Another strategic action that can exploit the installed-base effect is predatory pricing. An incumbent monopolist supplying the old technology may face a threat of competitive entry by a new technology. One weapon he can deploy against this is to engage in temporary price reductions, thus keeping the competition at bay, until his proprietary technology has so large an installed base that incompatible competitive entry becomes impossible. Because of the installed-base effect, predation can prevent future entry.

A Model without New Users. In our second model, we suppose that there are no new users: the network on the new technology must be built up through the switching of users from the old technology. By switching, a user encourages others to switch in two ways: the new network is made more attractive; the old less so. In our model, we assume that there are just two users, and that once one has switched the other will surely want to follow.

Where the users are firms, each firm's eagerness to switch will depend, inter alia, on the condition of its current equipment (does it need replacing anyway?), how well its current product line is doing (does the firm have other reasons to change the products it is offering?), whether it is already planning to add new capacity, etc. Looking forward the firm will thus be uncertain about when it will be advantageous for it (and for its rivals) to switch. We model this uncertainty about unpredictable eagerness to switch by supposing that "opportunities to switch" for each firm arrive randomly over time. (In particular, we assume that these opportunities follow a Poisson arrival process.) When such an opportunity occurs, the firm has a choice: it can switch now, or it can wait until its rival has an opportunity to switch, hope its rival switches at that opportunity, and then switch itself at *its* next opportunity.

In equilibrium, the firms may switch too reluctantly (excess inertia), too eagerly (excess momentum), or efficiently. The possible inefficiencies arise from two externalities in the model. First, when a firm switches, its rival loses some network benefits while they are using incompatible technologies, and the switching firm ignores this in its calculations.[8] Second, even if users unanimously favor a switch, each user may prefer the other to switch first.[9] As a result, switching may be delayed.

Adoption decisions will in general depend on prices. However, since we focus on the effects of the network externalities, we treat price as exogenous and suppress it. This is appropriate if the technologies are competitively supplied: W. Brian Arthur (1983) calls this "unsponsored technology." Michael Katz and Carl Shapiro (1986) discuss the effects of sponsorship. We consider price explicitly only when it has a strategic role, namely in our discussion of preannouncements and predation.

The paper is organized as follows: Section I presents the model with new users. Section II presents the model with no new users and a Poisson arrival process of switching opportunities.

I. A Model with New Users

In this section we study a model in which potential users arrive over time. Before time

---

[8] Indeed, this harm to the rival may provide an added incentive to switch if structural conditions are such that predation is possible.

[9] We call this the *penguin effect*. Penguins who must enter the water to find food often delay doing so because they fear the presence of predators. Each would prefer some other penguin to test the waters first.

T^*, only technology U is available. At T^*, a new technology, V, becomes available. Until we discuss preannouncements, we assume that the new technology is unanticipated. Those who adopted U before T^* do not switch, while subsequent users can choose which technology to adopt.

For simplicity, we suppose that users are infinitesimal and arrive continuously over time with arrival rate $n(t) \geq 0$ (i.e., we do not study shrinking markets). Write $N(t) = \int_0^t n(t')dt'$. Denote by $u(x)$ a user's flow of benefits from technology U when the size of the U network is x (i.e., when the set of users of technology U has measure x). The presence of network externalities implies that $u'(x) > 0$. For reasons that will become clear when we discuss equilibrium, we focus on two extreme possibilities. If everyone from T^* onwards adopts the new technology V, we call this the *adoption* outcome. If nobody adopts V, we call it *nonadoption*.

In the case of nonadoption of V, a user who adopts U at time T gets a present-value payoff of

$$(1) \quad \bar{u}(T) \equiv \int_T^\infty u(N(t))e^{-r(t-T)}dt,$$

where r is his or her discount rate.[10]

Periodically we will construct examples using the linear case:

$n(t)=1;\ N(t)=t$

$u(x)=a+bx$ (so $u(t)=a+bt$).

Here a represents the "network-independent" benefits: the value to a user if there are no other users in the network. At time t, the network has grown to size t, giving rise to network-generated benefits of bt. For this case, we have

$$(2) \quad \bar{u}(T) = \int_T^\infty (a+bt)e^{-r(t-T)}dt = \frac{a}{r} + b\int_0^\infty (\tau + T)e^{-r\tau}d\tau = (a+bT)/r + (b/r^2).$$

This has an appealing interpretation. The user who adopts the technology at time T join a network of size T, and so gets an initial benefit flow $(a+bT)$. The net present value of benefits if the network size remains unchanged is $(a+bT)/r$, which is the first term of the expression. Naturally, this is increasing in a, b, and T and decreasing in r. The second term, b/r^2, gives the benefit to the adopter at time T from the *future* growth in the network.[11] This term is increasing in b, the rate of network growth.

Similarly, we define the net present value of benefits to a user who adopts the old technology at time T *and is the last user to adopt it*. This is

$$\tilde{u}(T) \equiv u(N(T))\int_T^\infty e^{-r(t-T)}dt = u(N(T))/r.$$

In the linear case this is $(a+bT)/r$.

We define similar functions for the new technology V. Note, however, that when V is introduced, U already has an installed base of size $N(T^*)$. Therefore, if all new adopters after time T^* use V (and no current users of U switch), the benefit flow at time $t \geq T^*$ from using V is $v(N(T) - N(T^*))$. The analogous expression to (1) is

$$(3) \quad \bar{v}(T) = \int_T^\infty v(N(t) - N(T^*)) \times e^{-r(t-T)}dt \quad \text{for } T \geq T^*.$$

---

[10] Consumers' surplus is thus $\bar{u}(T)$ minus the price of technology U at time T. If we assume that price is constant over time and across technologies, then prices are common to most comparisons and can be ignored. While this is not realistic, it considerably simplifies notation and involves no loss of insight. The only comparison for which price is relevant is when the owner of a U machine considers scrapping it to buy a V machine. We assume that the price is high enough that this will not happen.

[11] Each unit of time the network grows by b. The net present value of benefits from this increase in the network size (evaluated at the time the growth occurs) is b/r. Thus the growth gives rise to a benefit stream of b/r per unit of time, which has a present value of $(b/r)/r = b/r^2$. It is this component of $\bar{u}(T)$ that is in jeopardy from the introduction of new technology.

In the linear case we suppose that

$$v(x) = c + dx = c + d(t - T^*).$$

Thus V can be superior to U in either of two ways: It can offer a higher *network-independent* benefit ($c > a$), or it can have higher *network-generated* benefits ($d > b$). In this linear case we have

$$\bar{v}(T) = \int_T^{\infty} [c + d(t - T^*)] e^{-r(t-T)} dt$$

$$= \frac{c + d(T - T^*)}{r} + \frac{d}{r^2} \text{ for } T \geq T^*.$$

The interpretation here is the same as for (2). Also $\tilde{v}(T)$ is given by the first term of $\bar{v}(T)$ as before.

A. *Equilibrium*

A strategy for the user who arrives at time $t \geq T^*$ consists of a choice of U or V as a function of what has happened before t, that is, the choices of those users who arrived between T^* and t. A Nash equilibrium specifies a strategy for each player that is optimal for him given the strategies of others. We now examine the Nash equilibria of the game and ask whether they are efficient.

In this game, there is a *bandwagon effect*: if a set of users adopts one technology, then that same choice thereby becomes more attractive to all other users. Accordingly, we focus on the two extreme possible outcomes: adoption and nonadoption. (Except for knife-edge cases, these are the only equilibria.)

Adoption is a subgame-perfect Nash equilibrium if $\bar{v}(T^*) \geq \tilde{u}(T^*)$. For then, if the first user to choose expects everyone else to adopt V, it is optimal for him to do likewise; and subsequent adopters then find the comparison all the more favorable to V. Moreover, because users are infinitesimal, any deviation from equilibrium by a single user will not affect the calculations of subsequent users, so the Nash equilibrium is indeed subgame perfect.[12]

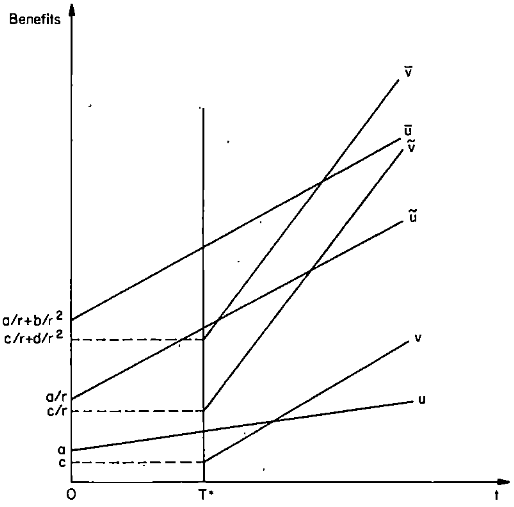


Figure 1: The Linear Example

If $\tilde{v}(T^*) < \tilde{u}(T^*)$, on the other hand, then it is a dominant strategy for all users close enough to T^* to choose U. Once they have done so, the installed base on U is all the larger and so later users will also adopt U. Figure 1 illustrates this in the linear case with $d > b$ and $a > c > 0$. When a few users just after T^* adopt U, the $\bar{v}(t)$ curve shifts horizontally to the right. Consequently, $\tilde{u}(T) - \bar{v}(T)$ is even larger for $T > T^*$ than at T^*. In this case, then, nonadoption is the unique equilibrium.

Similarly, if $\bar{u}(T^*) \geq \tilde{v}(T^*)$ then nonadoption is a (subgame-perfect) Nash equilibrium, because users near T^* will prefer U if they expect later users to adopt U, and later users will then find U all the more attractive. If $\bar{u}(T^*) < \tilde{v}(T^*)$, then it is a dominant strategy for early choosers to choose V, and once the bandwagon is rolling this dominance can only be strengthened: adoption is the unique equilibrium.[13]

---

[12] If we had a finite set of users, it could be that this Nash equilibrium is not perfect. If users are not infinitesimal, they may be able to affect later users' decisions. For a discussion of this strategic bandwagon power, see our earlier paper (1985).

[13] There is another case in which adoption is the unique equilibrium. Suppose that users are discrete rather than infinitesimal, and that (i) $\bar{v}(t) > \bar{u}(t)$ for all t, and (ii) $\tilde{v}(t) > \bar{u}(t)$ for all $t \geq T'$ for some T'. Then if all users between T^* and T' adopt V, it becomes a dominant strategy for later users to adopt V. Knowing this, it is optimal for the last user before T' to choose V if all previous users did so. Continuing by backwards induction in the usual way, we see that adoption is the unique perfect equilibrium. This is similar to Proposition 1 of our earlier paper (1985).

Since generally $\bar{u}(T^*) > \tilde{u}(T^*)$ and $\bar{v}(T^*) > \tilde{v}(T^*)$, at least one of the conditions, $\bar{u}(T^*) \geq \tilde{v}(T^*)$ or $\bar{v}(T^*) \geq \tilde{u}(T^*)$, must hold. So an equilibrium always exists in our model. However, the conditions are not mutually exclusive. When they hold simultaneously, then both adoption and nonadoption are equilibria: because the network externality makes it desirable to do what others are going to do, users' expectations drive the equilibrium.

The above results can be summed up in

PROPOSITION 1: *Adoption is a perfect Nash equilibrium if and only if* $\bar{v}(T^*) \geq \tilde{u}(T^*)$. *Nonadoption is a perfect Nash equilibrium if and only if* $\bar{u}(T^*) \geq \tilde{v}(T^*)$. *At least one of these conditions must hold, so equilibrium exists. They may hold simultaneously, in which case there will be multiple equilibria. If just one holds, we have a unique equilibrium.*

B. *Efficiency*

There are both positive and negative welfare effects from the adoption of the new technology. If V is adopted, each user for whom $\bar{v}(T) > \bar{u}(T)$ (typically one who arrives well after T^*) gains $\bar{v}(T) - \bar{u}(T)$. However, there are two groups of losers. First, early adopters lose $\bar{u}(T) - \bar{v}(T)$ if they adopt V. Second, users who are stuck with old installed-base technology suffer a loss at the time that V is adopted. Evaluated at T^* that loss is equal to $\bar{u}(T^*) - \tilde{u}(T^*)$ for each user in the installed base. This is the present value of the loss to U users resulting from the fact that the U network ceases to grow after T^*. In the linear case that is equal to b/r^2 for each user, or a total of bT^*/r^2.

The present value of the net gain in welfare from the adoption of V is

$$G \equiv \int_{T^*}^{\infty} [\bar{v}(t) - \bar{u}(t)]\, e^{-r(t-T^*)} dt - bT^*/r^2.$$

The first term represents the gain (loss) to users who arrive after T^*, while the second term is the loss to the installed base. Rewriting,

$$\text{(4)} \quad G = \int_{T^*}^{\infty} \left[\frac{c + d(t - T^*)}{r} + \frac{d}{r^2} - \frac{a + bt}{r} - \frac{b}{r^2} \right] e^{-r(t-T^*)} dt - \frac{bT^*}{r^2}$$

$$= \int_{0}^{\infty} \left(\frac{d\tau}{r} - \frac{b}{r}(\tau + T^*) \right) e^{-r\tau} d\tau + \frac{d-b}{r^3} - \frac{bT^*}{r^2} + \frac{c-a}{r^2}$$

$$= \frac{2(d-b)}{r^3} - \frac{2bT^*}{r^2} + \frac{c-a}{r^2}$$

$$= [2(d-b) - 2rbT^* + r(c-a)]/r^3.$$

Thus $G > 0$ if and only if $2(d-b) - 2rbT^* + r(c-a) > 0$.

We now compare this efficiency condition with equilibrium conditions. Adoption is an equilibrium if $\bar{v}(T^*) \geq \tilde{u}(T^*)$, which in the linear case reduces to

$$\text{(5)} \qquad r(c-a) - rbT^* \geq -d.$$

Adoption is the unique equilibrium if $\tilde{v}(T^*) > \bar{u}(T^*)$, which reduces to

$$\text{(6)} \qquad r(c-a) - rbT^* > b.$$

From (4), (5), and (6) we see that $(c-a, bT^*)$ space can be divided as in Figure 2. Depending on how d compares with $b/2$ and with $2b$, we have three cases.

Case 1: $2(d-b) > d$ (i.e., $d > 2b$). The $G = 0$ curve intersects the vertical axis below $(-d/r)$. This creates a region (region A) in which switching is efficient but is not an equilibrium: clear-cut excess inertia. In this region, $(c-a)$ is sufficiently negative to put off early potential adopters of V, but is not so negative as to make a switch socially inefficient.

Case 2: $(-d/r) < -2(d-b)/r < b/r$ (i.e., $b/2 < d < 2b$). There is no unique-equilibrium excess inertia (region A vanishes), but it is possible that nonadoption is

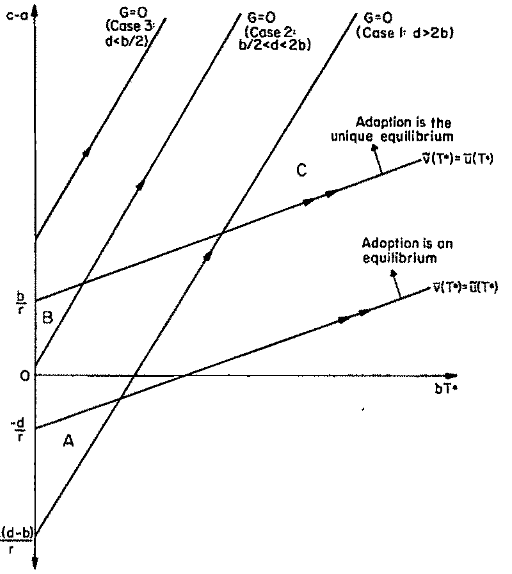


FIGURE 2: EFFICIENCY AND EQUILIBRIUM IN THE LINEAR CASE

an equilibrium even though adoption would be efficient. This occurs in region *B*.

Case 3: $-2(d-b)/r > b/r$ (i.e., $d < b/2$). There is *never* excess inertia (even region *B* vanishes): if adoption is efficient, then it is the unique equilibrium.

We also see from Figure 2 that there is always the possibility of excess momentum (region *C*): adoption is inefficient but is the unique equilibrium. This occurs when both bT^* (the installed base) and $(c-a)$ (the network-independent advantage of the new technology) are large. Then, V is adopted, and the installed base is inefficiently "stranded."

PROPOSITION 2: *In the linear case,*

(*i*) *If* $d > 2b$, *then there is a region in which adoption would be efficient but is not an equilibrium. There is excess inertia.*

(*ii*) *If* $b/2 < d < 2b$, *then adoption is an equilibrium wherever it is efficient. However, it need not be the unique equilibrium. There may be excess inertia.*

(*iii*) *If* $d < b/2$, *then adoption is the unique equilibrium whenever it is efficient. There cannot be excess inertia.*

(*iv*) *If the installed base and the network-independent superiority of* V *are both large, then there is unambiguous excess momentum: adoption is inefficient but it is the unique equilibrium.*

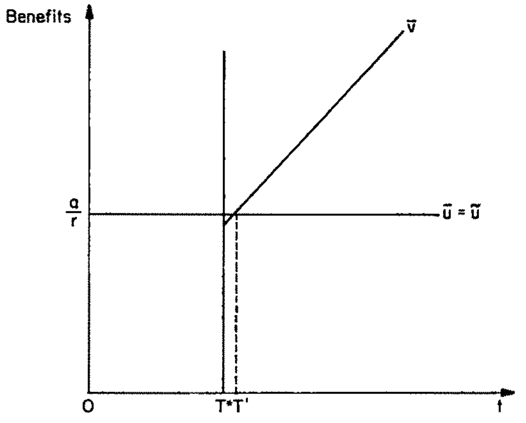


FIGURE 3: AN EXAMPLE OF EXCESS INERTIA

Although this analysis applies only to the linear case, it is instructive. It suggests that clear-cut excess inertia (inefficient nonadoption being the unique equilibrium) may be exceptional,[14,15] but it can arise. An extreme case is presented in Figure 3. Here there are no network benefits to the existing technology: $\bar{u} \equiv \tilde{u}$. Furthermore, the first potential user of the new technology V has only a slight preference for the old technology U ($\bar{v}(T^*) < \tilde{u}(T^*)$). However, even this slight preference causes him to adopt the old technology. All later users face the same

---

[14] The intuition for this result is the following. For it to be certain that the innovation will fail, it must be the case that $\bar{v}(T^*) < \tilde{u}(T^*)$, i.e., $rbT^* > d + r(c-a)$. This means that d cannot be too large. This can be seen by referring to Figure 1: as d grows so does $\bar{v}(T^*)$, eventually leaving the range where nonadoption is the unique equilibrium. However if d is small, the advantage of the new technology over the old is reaped only in the distant future, where discounting renders it less valuable. Of course, a higher value of d is possible (while still maintaining the inequality $\bar{v}(T^*) < \tilde{u}(T^*)$) if either r or T^* is large. However, if T^* is large, the loss to the installed base is also large, while if r is large, the benefits from the new technology (which are reaped only in the future) are also correspondingly less valuable.

[15] In our earlier paper (1986), we reach a similar conclusion in a static model in which users have different preferences over U and V.

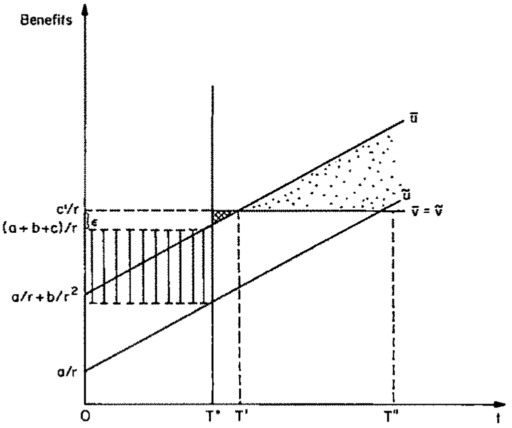


FIGURE 4: AN EXAMPLE OF EXCESS MOMENTUM

choice and make the same decision. Thus, *V* is not adopted although it would clearly be welfare enhancing.

"Excess momentum" can also arise. For example, consider the linear case and suppose that $c \equiv [a + bT^* + (b/r) + r\varepsilon]$ and $d = 0$. In that case $\bar{v}(t) = \tilde{v}(t) = c/r$, $\bar{u}(T^*) = ((a + bT^*)/r) + (b/r^2)$ and $\tilde{v}(T^*) - \bar{u}(T^*) = \varepsilon$. Thus for any $\varepsilon > 0$, the unique perfect equilibrium is that *V* is adopted. The situation is illustrated in Figure 4. The (undiscounted) gain to adopters from adoption of *V* is the shaded triangle (the discounted gain is even smaller). The loss to the installed base from the adoption is the shaded rectangle. There is also a loss to later adopters (dotted region). Since the shaded triangle can be made arbitrarily small (by shrinking ε) there obviously can be excess momentum. Indeed, it is possible that the losses are quite large and the gains minuscule, and yet *V* is adopted!

C. *Anticompetitive Behavior: Product Preannouncements and Predatory Pricing*

1. *Product Preannouncements.* So far, we have assumed that the new technology *V* becomes available unexpectedly. However, if potential users learn in advance (via a "product preannouncement," say) that *V* will become available, they may wait for it rather than adopt *U* immediately. But then *V*'s installed base will start off much larger than otherwise. It is therefore possible that a technology *V* will be adopted with such a product preannouncement that would not have been adopted otherwise. Without the preannouncement, the old technology might have developed an unstoppable momentum, whereas the preannouncement can prevent this "bandwagon effect."

Below, we provide conditions under which nonadoption is the unique equilibrium with no preannouncement, but adoption is an equilibrium with the preannouncement. Furthermore, if we modify the model slightly so that users are discrete rather than infinitesimal, then adoption is the *unique* equilibrium with the preannouncement. We also give an example in which the preannouncement reduces welfare.

Suppose that a preannouncement is made at $T^* - \tau$ for a new technology *V* that will become available at T^*. Assume that users in the interval $[T^* - \tau, T^*]$ who wish to do so can prepurchase technology *V*. (In practice, popular new products often have delivery lags and it is desirable to order in advance.) For simplicity, we assume that $n(t)$ is constant. Then if all potential users in the interval $[T^* - \tau, T^*]$ adopt *V*, and all later users also adopt *V*, the net present value of adopting *V* at $t \geq T^*$ is $\bar{v}(t + \tau)$. (This is the value it would have been at $T^* + \tau$ with adoption but without the preannouncement.) We can show

PROPOSITION 3: *If (i) $\bar{v}(T^*) < \tilde{u}(T^*)$ and (ii) $\tilde{v}(T^* + \tau - t')e^{-rt'} > \bar{u}(T^* - t')$ for all $t' \in [0, \tau]$, then*

(a) *Without the preannouncement, nonadoption is the unique equilibrium;*

(b) *With the preannouncement, adoption is an equilibrium.*

(c) *If, in addition, users are discrete, then with the preannouncement, adoption is the unique equilibrium.*

Assumptions (*i*) and (*ii*) are illustrated in Figure 5. When all the potential users between $(T^* - \tau)$ and T^* adopt *V*, the $\bar{v}$ and $\tilde{v}$ curves are "shifted to the left" by τ. The outlook for a potential user at $T^* + t$ absent the preannouncement is the same as the outlook of a potential user at $T^* + \tau + t$ with the preannouncement.

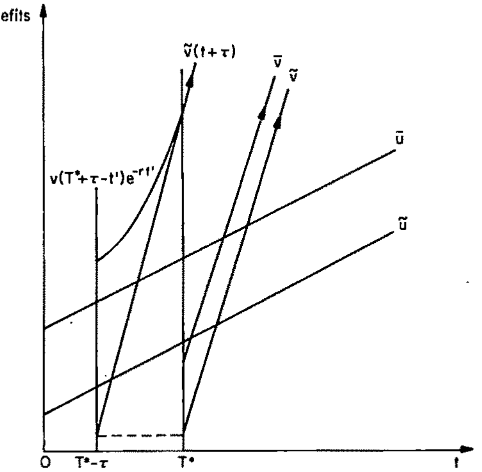


FIGURE 5: THE EFFECT OF A PREANNOUNCEMENT ON ADOPTION

PROOF OF PROPOSITION 3:

Assumption (*i*) implies (a) and assumption (*ii*) implies (b), both by Proposition 1. The proof of (c) uses the backwards induction argument of Proposition 1 of our earlier paper (1985) (see also fn. 2 above). Briefly, the argument is as follows: If all users from $T^*-\tau$ to T^* adopt V, then so will all later users. So consider the user who arrives "just before" T^*. If all prior users have adopted V, then so will he (since V will then be adopted by all later users). The backwards induction argument now applies the same reasoning to the previous arrival, and so on backwards.

Thus preannouncements can be powerful. We now show that they may be harmful.

PROPOSITION 4: *The preannouncement may reduce welfare, even though the conditions of Proposition* 3 *hold.*

PROOF:

We construct a linear example as follows: First, assume that $b<r$. Let $\tau \equiv \inf\{t'$: $\bar{u}(T^*-t')<\tilde{u}(T^*)-\varepsilon\}$ for some (arbitrarily small) ε. Second, let

$$v(x)=\begin{cases}0 & \text{if} \quad x\le\tau;\\ c' & \text{if} \quad x>\tau\end{cases}$$

(The new technology is valueless if the network size is less than τ and generates a benefit flow c' otherwise.) Now set c' so that $(c'/r)e^{-r\tau}>\bar{u}(T^*-\tau)$. Then $(c'/r)e^{-rt'}>\bar{u}(T^*-t')$ for all $t'\le\tau$. Thus the conditions of Proposition 3 are satisfied. Notice that because u is linear, this construction can be carried out for any T^*. One can check that the welfare gain (evaluated at T^*) to adopters of the new technology is bounded above and is independent of T^*. However, the welfare loss to the installed base is $T^*[\bar{u}(T^*-\tau)-\tilde{u}(T^*-\tau)]=T^*b/r^2$, which can be made arbitrarily large by increasing T^*.

We have argued that a preannouncement may reduce welfare. Dominant firms have often faced antitrust charges of "anticompetitive" product preannouncements. (See fn. 6 above.) In the above analysis we have made pricing exogenous, and so our model does not yet show that the welfare-reducing preannouncements can be anticompetitive. However, it is a simple matter to complete the argument.

Suppose that the old technology U is provided competitively and that the u function represents consumers' benefits, given competitive pricing. Suppose further that the new technology V is provided by a monopolist and that the v function represents the benefit to consumers when the monopolist prices so as to maximize profits assuming that his proprietary V technology will be adopted. Under the assumptions of the example, absent a preannouncement, the monopolist's product will fail to be adopted. Of course, the monopolist might be able to induce adoption by offering a discount until V has become widely enough adopted. However, the preannouncement achieves the same result costlessly. Competition is destroyed and welfare may also be reduced.

The preannouncement is likely to be most effective where the current technology could otherwise greatly increase its network value in a short time between announcement time and introduction. Then the preemptive effect of the preannouncement will be crucial. So, especially when targeted against a fledgling technology, the preannouncement may well be anticompetitive.

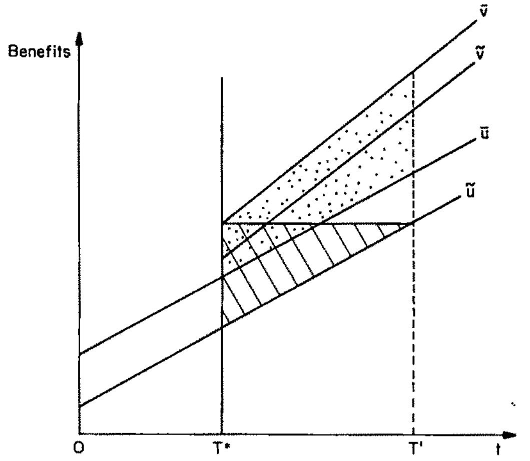


FIGURE 6: THE COST OF PREDATION

2. *Predatory and Strategic Pricing*. We consider the case in which the existing technology is provided by a monopolistic incumbent. We show below that if technology V is introduced, it may be both feasible and profitable for the incumbent to prevent the adoption of V by temporarily charging less than the monopoly price. By doing so, it can continue to build up its installed base and achieve an insurmountable lead, whereas absent such strategic pricing, V would be adopted. This strategic pricing is analogous to "penetration pricing" as analyzed by Katz and Shapiro (1986).

Suppose that when the incumbent sets its profit-maximizing prices over time (assuming no new technology), the flow of net benefits to users is given by the u function. Suppose further that the new technology V will be competitively supplied and that the v function represents benefits net of the competitive price. Finally, suppose that u and v are as shown in Figure 6.

Since $\tilde{v}(T^*) > \bar{u}(T^*)$, adoption is the unique equilibrium absent strategic pricing. But the incumbent can induce the user who arrives at T^* to stay with U by reducing the price of U by (just over) $\bar{v}(T^*) - \tilde{u}(T^*)$. Then that user would prefer being the last user on U to being the first user of V, even if all later users adopt V. If the incumbent provides such a discount to all the users who arrive between T^* and T', it will then become impossible for V to enter (since $\bar{v}(T') < \tilde{u}(T')$), *even if the incumbent then prices without regard to entry.*

The (undiscounted) cost to the incumbent of this strategy is given by the shaded area. Depending on the parameters of the model, the benefits may well exceed the costs, although in the short run the monopolist sacrifices profits. The private costs of pricing strategically are not the same as the social costs. The cost to the monopolist occurs through price reductions which are merely transfers to users. The social costs are the costs to users who arrive after T^* from the fact that the new technology is not adopted. The social benefit is that the installed base is protected against "stranding." Thus the welfare effects are ambiguous.

The sole purpose of the incumbent's pricing policy is to drive its rivals out of the industry and it may involve a short-run sacrifice of profits. Moreover, the monopolist can raise its price once the rivals have left the industry, without inducing reentry, even if there are no entry or exit costs. This is because if the incumbent fights off entry until T', it thereafter has an insurmountable advantage of installed base. If, in addition, welfare is reduced by the action, this is a classic case of predatory pricing.

Two leading tests for diagnosing predatory pricing may miss this kind of predation, however. The test proposed by Philip Areeda and Donald Turner (1975), which diagnoses predation if the firm prices below average variable cost, can yield a false negative: if the network benefits of installed base are sufficient, this kind of predation does not require pricing below the incumbent's, *or the entrants'*, average variable cost.

Janusz Ordover and Robert Willig's (1981) test asks whether the firm's action would have been optimal if the entrant faced no reentry costs. If so, then the action is deemed not to be predatory. Reentry costs are defined as "the cost that a firm that has exited a market must incur to resume production" (p. 12). Here, however, "reentry costs" are not the point. The opportunity to supplant the monopolist is only available until time T'. Through its predatory action, however, the incumbent closes this window of opportunity.[16] This evanescence of opportunity is assumed away in the Ordover-Willig test. Thus this test may also yield a false negative.

Thus we see that, on the one hand, standard tests may miss this kind of predation. On the other hand, if the stranding externality is important enough, the strategic pricing may be welfare enhancing. Because of this, it would be extremely difficult (if not impossible) to frame a legal rule that would make the correct diagnosis in all cases. What our analysis suggests, however, is that a greater than usual degree of caution is required in cases where "a window of opportunity" is important.

II. A Model without New Users

In this model, two agents, 1 and 2, are initially using a preestablished standard or technology, U. Another technology, V, is available; however, each agent has only occasional chances to switch. To be precise, each agent encounters switching opportunities randomly in a Poisson process with rate $\lambda > 0$.

An agent with a chance to switch faces the following choice. If he (or she) switches, then he knows that the other agent will not immediately be able to switch, so that they will be incompatible for a while. If he does not switch, then the other may or may not choose to switch at his next opportunity. We will consider agents' private incentives to switch or not to switch, and the nature of the externalities involved.

We write $u(k)$ for each agent's flow of net payoff if he is on standard U alone ($k=1$), or together with the other agent ($k=2$). Likewise, we define $v(k)$. Notice that this assumes that the firms' payoffs are symmetric.

As above, we interpret the two "agents" as *users* of the technology. If one of them switches, they become incompatible (because the other cannot immediately switch). In this interpretation, "installed base" is the nonswitcher. It is natural to assume that network externalities are positive:[17]

$$u(1) < u(2) \tag{7}$$

$$v(1) < v(2). \tag{8}$$

For notational convenience, we normalize $u(2)$ at 0.

A. *Efficiency*

It should be clear that the efficient rule is either never to switch, or to switch as soon as possible. Which is better?

Suppose that agent 1 has a switching opportunity at a time we call $t=0$. Let $\tilde{t}$ be the (random) next switching opportunity for agent 2. Then the social value of switching is

$$\int_0^{\tilde{t}}[v(1)+u(1)]\,e^{-rt}dt+\int_{\tilde{t}}^{\infty}2v(2)e^{-rt}dt \tag{9}$$

$$= [v(1)+u(1)]\big((1-e^{-r\tilde{t}})/r\big) + 2v(2)\big(e^{-r\tilde{t}}/r\big). \tag{10}$$

It is well-known that

$$E(e^{-r\tilde{t}}) = \lambda/(r+\lambda). \tag{11}$$

---

[16] As time passes and the installed base grows, the barriers to entry rise and the incumbent becomes more protected. In our setting this is due to the benefits from compatibility; more generally, the same will be true where brand recognition or a product's reputation are important in determining consumer choice. Learning by doing will have a similar effect.

[17] In a second interpretation, our two agents are competing *suppliers* of a durable good embodying technology U or V. Buyers are concerned with network externalities: in particular, with compatibility with the installed base of equipment. If this network effect is strong enough, it may be the case that

$$u(1) > u(2). \tag{a}$$

Also, if the installed-base effect is *not* so strong, so that buyers begin to buy the new technology as soon as it is available, there may be a *first-mover advantage*:

$$v(1) > v(2). \tag{b}$$

Both (a) and (b) are influenced by buyers' expectations of whether the new technology will succeed, and by the relative sizes of the two suppliers.

Thus the expected social value of switching is

$$(12)\quad \frac{v(1)+u(1)}{r+\lambda}+\frac{2v(2)(\lambda/r)}{r+\lambda}.$$

In other words, we get

PROPOSITION 5: *Switching is socially efficient if and only if*

$$(13)\qquad v(1)+u(1)+\lambda\frac{2v(2)}{r}\geq 0.$$

Equation (13) has a straightforward interpretation. Suppose a firm switches. The net social benefit has two components. The immediate change in benefit flow is $v(1)+u(1)-2u(2)$, or $v(1)+u(1)$. In addition, there is an instantaneous probability λ of achieving a joint switch which is worth $v(2)/r$. The qualitative properties of (13) are not surprising. For example, if $v(2)<0$ then, by assumption, $v(1)<0$. Since in any case $u(1)<0$, switching is not efficient. If $v(2)>0$ then switching is more desirable as λ increases (less time is spent incompatibly in transition) or as r decreases (transition effects are less important).

B. *Equilibrium*

As in Section I, our chief concern is to investigate when equilibrium will be efficient, when there will be excess inertia, and when there will be excess momentum.

We ask first in what circumstances a user will be willing to switch first. As we will see, the answer depends on whether he expects that the other would be willing to switch first.

LEMMA 1: *A user will be prepared to switch first when he believes that the other user would switch first, if and only if*

$$(14)\quad (\lambda+r)v(1)+\lambda[v(2)-u(1)]\geq 0.$$

PROOF:

Since we assume $v(2)>u(1)$, a user who switches first will always be followed as soon as possible. Therefore we can use the method of (10)–(12) to show that the payoff to switching first is

$$(15)\quad (r+\lambda)^{-1}[v(1)+(\lambda/r)v(2)].$$

The payoff to waiting is a flow of 0 until the other user switches, then a flow of $u(1)$ until it is possible to follow. The value of this is

$$(16)\quad \frac{\lambda}{(r+\lambda)}\frac{1}{(r+\lambda)}[u(1)+(\lambda/r)v(2)].$$

Comparing (15) and (16) yields (14).

LEMMA 2: *When a user believes that the other will never switch first, he will switch first, if and only if*

$$(17)\qquad rv(1)+\lambda v(2)\geq 0.$$

PROOF:

If he does not switch, he gets zero forever. If he does, he gets $v(1)$ until the other user follows. By the method of (10)–(12), the present value of this is

$$\frac{v(1)}{r}\frac{r}{r+\lambda}+\frac{v(2)}{r}\frac{\lambda}{r+\lambda}.$$

Condition (17) follows immediately.

LEMMA 3: *If both* (14) *and* (17) *hold, then a user will switch first whatever his beliefs about whether the other would do so. If neither* (14) *nor* (17) *holds, then a user will never switch first, whatever his beliefs about the other's willingness to do so.*

PROOF:

Suppose that agent 1 regards the time at which agent 2 will switch (if 1 does not switch first) as a random variable $\tilde{s}$. We write $\tilde{s}=\infty$ if 2 never switches first. If 1 does not switch first, then he gets a flow $u(2)\equiv 0$ until the random time $\tilde{s}$, at which point his present discounted value becomes

$$(18)\quad E\left\{\int_0^{\tilde{t}}u(1)e^{-rt}dt+\int_{\tilde{t}}^{\infty}v(2)e^{-rt}dt\right\}$$
$$=(r+\lambda)^{-1}(u(1)+(\lambda/r)v(2))$$

where $\tilde{t}$ is the (random) elapsed time after $\tilde{s}$ before agent 1 has his next switching opportunity. Thus, at time 0, the expected present value of the "wait" strategy is

$$(19)\quad E[e^{-r\tilde{s}}](r+\lambda)^{-1}(u(1)+(\lambda/r)v(2)).$$

Finally, we note that $\tilde{s}$ cannot occur before agent 2's first switching opportunity, so that, by (11),

$$(20)\qquad 0 \le E[e^{-r\tilde{s}}] \le \lambda/(r+\lambda).$$

Using (20) and comparing (15) and (18), we see that (14) and (17) together make switching optimal whatever a user's expectations. The proof of the other half of the lemma is similar.

From the lemmas, we see that (given r and λ) the space of possible values of $u(1)/v(2)$ and $v(1)/v(2)$ divides into four regions, according to whether (14) and/or (17) holds. Adding the efficiency condition (13) to the picture yields Figure 7. We label the resulting six regions of the parameter space A to F. Switching is efficient in regions A, B, C; inefficient in regions D, E, F. What does equilibrium look like?

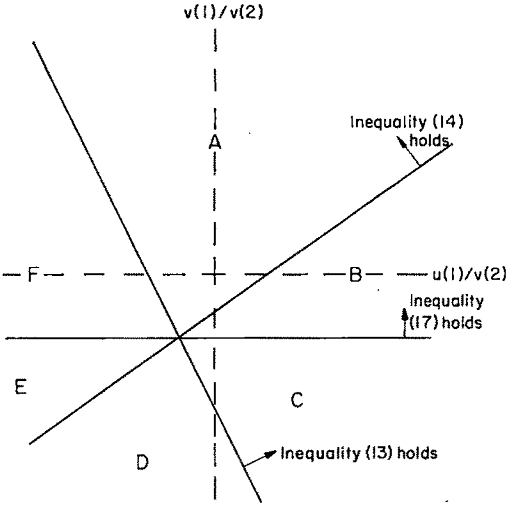


FIGURE 7: EFFICIENCY AND EQUILIBRIUM

In region A, (14) and (17) both hold. By Lemma 3, a user will switch as soon as possible, *whatever* he thinks the other would do (switch first, or wait). In this region, therefore, there is a unique (and efficient) perfect equilibrium: switching occurs as soon as possible.

In region B, (17) holds but (14) fails. This means that a user will switch first if he thinks the other would not, but would wait if he thinks the other would switch first. In this region, therefore, there are three perfect equilibria. In the first, user 1 is expected to (and will) switch as soon as he can, and user 2 will only follow. In the second, the roles are reversed. In the third equilibrium, each user plays a mixed strategy: he is just likely enough to switch first to make the other indifferent between switching first and waiting.

In each of these equilibria, switching occurs eventually, but is (on average) delayed. Fortunately, if r is large relative to λ (the case where delay matters), region B is small: (14) and (17) almost coincide.

In region C, both (14) and (17) fail. This means that neither user will switch first, whatever his beliefs about the other's willingness to switch first. Therefore switching never occurs, although it would be efficient to switch. This is the clearest kind of excess inertia. From (14) and (17), we see that this outcome is most likely when $v(1)$ is small and $u(1)$ is not too small.

In region D, (14) and (17) both fail: there will be no switching. Region D differs from region C in that this outcome is efficient in D.

In region E, (14) holds but (17) fails. So a user will switch first if and only if he thinks the other would do so. There are two pure-strategy equilibria. In one, there is no switching, and this is efficient. In the other, each user will switch at his first chance. However, this is only because if he does not switch he believes he will have to follow the other's switch. Such a *preemption equilibrium*[18] can arise even if $v(2)$, as well as $u(1)$ and $v(1)$, is negative.

---

[18] If switching opportunities are observable, we can argue against the preemption equilibrium as follows. If 2 expects 1 to switch first, but observes him refraining from doing so, we might expect him to infer that 1 will not switch first, rather than inferring that 1 will *now*

In region F, (14) and (17) both hold: switching will occur as soon as possible, although it is inefficient. This is a clear case of excess momentum.

What causes these inefficiencies? The fact that there will be a period of incompatibility is not itself the source of an externality, but the difference between $v(1)$ and $u(1)$ is. In switching, a user calculates whether it is worth suffering $v(1)$ for a while in order to achieve $v(2)$ later; for efficiency, he should ask about suffering $[v(1)+u(1)]/2$. If $v(1) > u(1)$, then a user will be too ready to initiate a switch; if $v(1) < u(1)$, he will be too reluctant.

Second, there is the penguin effect (see fn. 9). Even though the users might *both* wish for a joint switch *whoever* goes first, it may still be that each would prefer the other to go first. This leads to the kind of excess inertia that occurs in region B. Conversely, it is possible that neither user wants a switch, but each prefers to switch first rather than second: this leads to preemption equilibrium.

III. Conclusions and Possible Extensions

There can be excess inertia, even with complete information, when we allow for the presence of an installed base. There are two externalities in a user's adoption decision: the stranding effect on the installed base, and an effect on the options available to later adopters. Since our previous model (in our 1985 paper) is essentially "timeless," the installed-base externality is absent. In that setting, identical adopters always reach the efficient outcome. In the present paper, installed-base users are somewhat tied to the old technology; this creates a bias against the new (perhaps superior) technology. Another inefficiency arises from the penguin effect.

The biases we identify can be turned to anticompetitive uses. We analyzed two: anticompetitive product preannouncements, and predatory pricing. First, product preannouncements may prevent a bandwagon from gaining momentum. Second, an incumbent firm may be able to deter entry by a credible threat of temporary price cuts in response to entry. This is so even when there are no reentry costs.

There are a number of ways in which the models could be generalized and extended. In the model with new users, we assumed that users live forever and products do not depreciate. Realistically, an installed base, once stranded, will shrink. The welfare cost of excess momentum will then be lower than in our model.

We could also examine the effect of a nonconstant arrival rate of new users. We would expect that if there is an unusual upturn in demand, perhaps because of a baby boom or an economic recovery, an innovation could gain a substantial network relatively quickly. Innovations may therefore be concentrated in such periods. Similarly, the destruction of installed base by war may clear the decks for innovation.

In the second model, the most interesting extensions would make the users asymmetric, either in the frequency with which they consider switching or in the value they attach to compatibility. Large firms plausibly care less about compatibility with small rivals than vice versa, and we therefore expect them to be de facto standard setters. Likewise, an agent who only rarely has a chance to switch may find his more flexible rival waiting for him to do so.

Another line of inquiry would consider multiple users in the second model. With equal-sized users, the first switcher sacrifices more in network benefits if there are many users. His strategic bandwagon power is also likely to be less. One might expect, therefore, that excess inertia will be a more serious problem in this case. This extension would form a close link with our earlier paper (1985). It would be equivalent to allowing stochastic decision lags, random order of

---

switch as soon as possible. If 1 believes that 2 would update as we suggest, then he can stop 2 from switching by refraining himself. Thus, for preemption to be a perfect equilibrium, updating after an unexpected failure to switch must take the form "I'm surprised you didn't switch; but I still expect you to switch at your next chance."

moves, and interim payoffs in that model. We believe that these are promising directions for further investigation.

REFERENCES

Areeda, Philip and Turner, Donald F., "Predatory Pricing and Related Practices Under Section 2 of the Sherman Act," *Harvard Law Review*, February 1975, *88*, 697–733.

Arthur, W. Brian, "On Competing Technologies and Historical Small Events: The Dynamics of Choice under Increasing Returns," mimeo., Stanford University, 1983.

David, Paul A., "Clio and the Economics of QWERTY," *American Economic Review Proceedings*, May 1985, *75*, 332–37.

Farrell, Joseph and Saloner, Garth, "Standardization, Compatibility, and Innovation," *Rand Journal of Economics*, Spring 1985, *16*, 70–83.

______ **and** ______, "Standardization and Variety," *Economics Letters*, January 1986, *20*, 71–74.

Fisher, Franklin M., McGowan, John and Greenwood, Joen, *Folded, Spindled and Mutilated: Economic Analysis and US vs IBM*, Cambridge: MIT Press, 1983.

Katz, Michael and Shapiro, Carl, "Network Externalities, Competition and Compatibility," *American Economic Review*, July 1985, *75*, 424–40.

______ **and** ______, "Technology Adoption in the Presence of Network Externalities," *Journal of Political Economy*, August 1986, *94*, 822–41.

Landis, Robin and Rolfe, Ronald, "Market Conduct Under Section 2—When is it Anticompetitive?," in F. M. Fisher, ed., *Antitrust and Regulation; Essays in Memory of John J. McGowan*, Cambridge: MIT Press, 1985.

Ordover, Janusz A. and Willig, Robert D., "An Economic Definition of Predation: Pricing and Product Innovation," *Yale Law Journal*, November 1981, *91*, 8–53.

Schelling, Thomas C., *The Strategy of Conflict*, Cambridge: Harvard University Press, 1960.

Oligopoly and Financial Structure: The Limited Liability Effect

By JAMES A. BRANDER AND TRACY R. LEWIS*

We argue that product markets and financial markets have important linkages. Assuming an oligopoly in which financial and output decisions follow in sequence, we show that limited liability may commit a leveraged firm to a more aggressive output stance. Because firms will have incentives to use financial structure to influence the output market, this demonstrates a new determinant of the debt-equity ratio.

The literature on financial structure and the literature on oligopoly have at least one common feature: they both place relatively little emphasis on the strategic relationships between financial decisions and output market decisions. In financial theory, the product market is typically assumed to offer an exogenous random return which is unaffected by the debt-equity positions of the firms in the market. Correspondingly, in the economic analysis of oligopoly, the firm's obligations to debt holders and the possibility of financial distress are usually ignored in modeling the strategic interaction between producers in the output market.

This approach of focusing separately on financial and output decisions is clearly useful in understanding certain aspects of both financial structure and strategic output market behavior. It seems equally clear, however, that there are important linkages between financial and output decisions.

The choice of financial structure can affect output markets in the following way, which we refer to as the limited liability effect of debt financing. As firms take on more debt, they will have an incentive to pursue output strategies that raise returns in good states and lower returns in bad states. The basic point is that shareholders will ignore reductions in returns in bankrupt states, since bondholders become the residual claimants. As debt levels change, the distribution of returns to shareholders over the different states changes, which in turn changes the output strategy favored by shareholders.[1]

A second possible linkage between output and financial markets is the strategic bankruptcy effect. Any one firm's susceptibility to financial distress depends on its financial structure, and its fortunes will usually improve if one or more of its rivals can be driven into financial distress. Therefore, firms might make output market decisions that raise the chances of driving their rivals into insolvency. Since the possibility of financial distress for each firm is contingent on its financial structure, this is a second channel for finances to affect output markets.

In this paper we examine the relationship between financial and output decisions in a formal structure capturing essential aspects of both modern financial and oligopoly theory. Our model focuses on the "limited liability" effect of debt financing. The strategic bankruptcy effects of financial decisions are analyzed in our companion paper (1985). The analysis we offer here illustrates two important points. First, output market behavior will, in general, be affected by financial structure. Second, foresighted firms will anticipate output market consequences of financial decisions; therefore, output market conditions will influence financial decisions.

---

*Faculty of Commerce and Business Administration, University of British Columbia, Vancouver, B.C. V6T 1Y8, Canada, and Department of Economics, University of California, Davis, CA 95616, respectively. In writing and revising this paper we have incurred substantial debts (without offering equity participation) to several people. We are grateful to three anonymous referees, and also thank Ron Giammarino, Robbie Jones, and Barbara Spencer for helpful comments. In addition we have received help from other colleagues at UBC in both Finance and Economics, and have benefited from presenting the paper in seminars at the Rand Corporation, UC-Berkeley, the University of Regina, the University of Saskatchewan, and the University of Southern California.

[1] The idea that limited liability creates a conflict of interest between bondholders and equity holders is described in Michael Jensen and William Meckling (1976) and in Stewart Myers (1977), and has been examined by Jeremy Bulow and John Shoven (1978), Jerry Green and Shoven (1983), and Varouj Aivazian and Jeffrey Callen (1980). These studies do not, however, explicitly consider the output market, which is the focus of this paper.

In the analysis to follow, we deliberately abstract from the physical capital investment decision by assuming that the firm's capital stock is fixed (at least temporarily). We adopt this approach so as to isolate the linkage between the firm's financial position and its behavior in the output market. If physical investment was allowed to vary endogenously with the firm's financial position, certain readily identifiable effects of financial decisions on output behavior would emerge. First, firms would have an incentive to undertake capital investments that lower the marginal cost of production in output markets so as to commit themselves to more aggressive positions in the output market. (This effect has been investigated by Avinash Dixit, 1980; Curtis Eaton and Richard Lipsey, 1980; and Brander and Barbara Spencer, 1983; among others.)

In addition, as emphasized by Michael Jensen and William Meckling and Stewart Myers (1977), the firm's debt-equity mix would affect the optimal investment strategy. A referee suggested that heavily leveraged firms might lower their level of physical capital investments in order to minimize the salvageable assets that would be lost to debtholders in the event of bankruptcy. This would in turn affect costs and the output market in the obvious way. These are interesting issues, but they are logically separate from the effects we identify; accordingly we abstract from the investment decision by assuming it is made before the debt-equity mix is determined. Output adjustments are therefore to be thought of as resulting from changes in variable inputs.

The basic model we investigate is a two-stage sequential duopoly game. In stage 1, the two firms decide upon financial structure. In stage 2, they select output levels taking as given the financial composition determined in stage 1. The equilibrium concept is the sequentially rational Nash equilibrium in debt levels and output levels. In other words, the second-stage outcome is a Cournot equilibrium in output which is correctly anticipated by the firms when choosing debt levels in the first stage.[2]

The output decisions of firms are made before the realization of a random variable reflecting variations in demand or costs. Once profits are determined, firms are obliged to pay debt claims out of operating profits, if possible. If profits are insufficient to meet debt obligations, the firm goes bankrupt and its assets are turned over to the bondholders. This simplified depiction of the relation between financial distress and output markets is not universally descriptive, but it does yield a tractable model that can be readily compared with similar models of output market rivalry in the oligopoly literature where financial structure is ignored.

Before proceeding it is important to place our analysis in context in both the financial and oligopoly literatures. A large part of the literature on financial structure can be regarded as a response to Franco Modigliani and Merton Miller (1958), who assert that, under certain conditions, the value of the firm is independent of financial structure. Our paper falls in the class of exceptions to the Modigliani-Miller Theorem. Following Alan Kraus and Robert Litzenberger (1973), among others, we appeal to a background of imperfect markets to allow departures from the Modigliani-Miller world. The exact nature of the market imperfections should become clear as we proceed. The standard treatment of choice of financial structure, as presented in Kraus and Litzenberger, involves a tradeoff between financial distress costs and the tax advantages of debt. Our model, which abstracts from both bankruptcy costs and taxes, points to the interaction between output and financial markets as a determinant of financial structure.

---

[2] In our context, sequential rationality is equivalent to what Reinhardt Selten (1975) has referred to as subgame perfection. The most important implication of subgame perfection is that players are restricted to credible threats: players cannot have equilibrium strategies that would call for them to carry out actions that would be against their best interests at the time the action is to be taken. This basic idea goes back at least to Thomas Schelling (1956).

Of the existing published literature in financial theory, our paper is closest in spirit to Sheridan Titman (1984) in which financial structure influences potential profits for suppliers of a durable good. Titman argues that low levels of debt constitute a commitment by a producer to stay in the market, which raises the value of the durable good because it means that the firm will be available to service the good in the future. (This effect operates as long as each firm has some advantage over rivals in servicing its own products.) In our model, financial structure serves as a commitment to a particular output strategy,[3] and is therefore related to the recent literature on commitment in oligopoly theory, in which physical capital, location, product choice, or *R&D* choices constitute a first stage which influences the Nash equilibrium in the subsequent output market. Standard references in this literature include Eaton-Lipsey and Dixit.[4] The equal opportunity duopoly structure we use is similar in form to Brander-Spencer.

An outline of the paper follows. Section I sets out the basic model. Section II is devoted to the output market equilibrium and, in particular, shows the dependence of output equilibrium on financial structure and compares the output equilibrium in the (base) case, in which equity holders control the firm, with the case in which debtholders control the firm. Section III examines the selection of debt levels and describes how output market considerations influence capital structure. Section IV contains concluding remarks and discusses extensions to the paper.

I. The Model

Firms 1 and 2 are rivals in an output market where they produce competing products q_1 and q_2, respectively. For concreteness, we assume there is Cournot quantity competition in the output market. Other forms of market rivalry involving advertising, *R&D*, or Bertrand price competition could be analyzed just as well using our model. The operating profit for firm i, which is defined as the difference between revenue and variable cost, is denoted by $R^i(q_i, q_j, z_i)$. The random variable z_i reflects the effects of an uncertain environment on the fortunes of firm i. It is assumed to be distributed over the interval $[\underline{z}, \bar{z}]$ according to density function $f(z_i)$. For simplicity we assume that z_i and z_j are independent and identically distributed.

We assume that R^i satisfies the usual properties: $R^i_{ii} < 0$, $R^i_j < 0$, and $R^i_{ij} < 0$ (where subscripts denote partial derivatives). We adopt the convention that high values of z_i lead to higher operating profits: $R^i_z > 0$, meaning that higher realizations of z_i correspond to better states of the world. The effect of z_i on marginal profit turns out to be very important in our analysis. We consider two possibilities.

(*i*) $R^i_{iz} > 0$. This corresponds to a situation where marginal profits are higher in better states of the world. This would arise, for example, if higher realizations of z corresponded to downward shifts of the marginal cost schedule, or to upward shifts in the marginal revenue schedule facing the firm. We take this to be the normal case.

(*ii*) $R^i_{iz} < 0$. This means that good states of the world are correlated with low marginal returns to extra sales. This case seems less likely to arise but it is possible.[5]

---

[3] We would also like to mention a Ph.D. thesis by Vojislav Maksimovic (1986) and a discussion paper by Franklin Allen (1985) which address the same fundamental issue as this paper: the strategic relationship between oligopoly and financial markets. Maksimovic represents simultaneous development of a modeling approach that is similar in some respects to ours and establishes some of the same insights. It also examines some interesting repeated game extensions of the basic model.

[4] See Eaton and Mukesh Eswaran (1984) for a very helpful synopsis of the strategic commitment paradigm applied to industrial organization.

[5] A rather contrived example can be constructed as follows. Suppose that firms i and j each have fixed

Another argument for considering the $R^i_{iz} < 0$ case arises when we assume that firms engage in other forms of competition besides quantity or price competition. For example, suppose that firms primarily compete through advertising, as is the case in some retail markets.[6] Let q_i be firm i's expenditure on advertising and assume that higher realizations of z_i correspond to cases where market demand is high. It seems plausible that, in some cases, when conditions are good, there might be little need for advertising, whereas advertising would be more effective in increasing sales in a market with sagging demand. In this case we would find that marginal returns from advertising, R^i_i, are inversely related to the state of nature, z_i, leading to the case $R^i_{iz} < 0$.

An alternative way of thinking of the sign condition on R^i_{iz} is to view q_i and z_i as separate "inputs" which go into making profits. The case $R^i_{iz} > 0$, which is the standard case under quantity competition, means that the two inputs are complementary, while the other case is the case of substitute inputs.

The financial structure of the firm is summarized by the variable D_i, which represents the debt obligation of firm i. As indicated in the introduction, we assume that the capital investment is made before the financing mix is decided upon. Choosing the debt level, with the total financing requirement fixed, would then fix the equity financing level by default. This interpretation is not necessary, however, and the firm could just as easily be viewed as giving the borrowed money directly to shareholders. Debt levels, in turn, are assumed to be chosen before output decisions are decided upon and are taken as given when output levels are chosen. Output decisions are then made before the uncertainty over demand or cost is resolved. It is not essential that there be no uncertainty resolved before output decisions are made. What is important is that there be some residual uncertainty left to be resolved after output decisions are made. Certainly this is true of most industries.

After production occurs and the uncertainty regarding firms' profits is settled, the firm is obliged to pay creditors D_i out of current profits. If the firm is unable to meet its debt obligations, its creditors are paid whatever operating profits are available.[7]

Given debt levels (D_1, D_2), the firm is assumed to choose output levels with the objective of maximizing the expected value of the firm to the shareholders.[8] This is what an owner-manager would choose to do, and is certainly what wealth-maximizing shareholders would want the firm to do. The value to the shareholders is referred to as the equity value and is represented by the letter V:

$$(1)\quad V^i(q_i, q_j;.) = \int_{\hat{z}_i}^{\bar{z}} \left(R^i(q_i, q_j, z_i) - D_i\right) f(z_i)\, dz_i,$$

where $\hat{z}_i$ is defined by

$$(2)\qquad R^i(q_i, q_j, \hat{z}_i) - D_i = 0,$$

assuming $\underline{z} < \hat{z}_i < \bar{z}$. When $z_i = \hat{z}_i$, firm i can just meet its debt obligations with nothing left over. The expression in (1) represents expected current-period profits net of debt obligations in good $(z_i \geq \hat{z}_i)$ states of the

---

production capacity k, and that they are rivals in a domestic market where they produce q_i and q_j, respectively, but they also sell their remaining outputs $k - q_i$ and $k - q_j$ in separated foreign markets where they do not compete with each other and where price is given by $p^i + z_i$. In this case, higher values of z_i are certainly good for the firm, but correspond to lower levels of marginal profit in the home market.

[6] See Richard Schmalensee (1976) for an analysis of oligopoly markets where firms compete in advertising. One could turn our model into a model of advertising by taking output as exogenously fixed, and interpreting q_i as advertising expenditure.

[7] For simplicity we assume that the asset value of the firm is zero, as if assets are completely used up in the production of output. Creditors can, therefore, collect only current operating profits if the firm becomes insolvent.

[8] The interesting possibilities that the rival firms may somehow be connected through interlocking directorships, or that they are both owned by a common group of shareholders are not considered here.

world. In bad states ($z_i < \hat{z}_i$), the firm earns zero as all of its earnings are paid to debtholders. (Note that we are assuming that shareholders of the firm are risk neutral with respect to the firm's returns and therefore have their interests served by maximization of expected equity value.)

Expression (2) shows the implicit dependence of $\hat{z}_i$ on D_i, q_i, and q_j. As this relationship is important in establishing the principal results of the paper, it is useful to report the following derivatives:

$$\text{(3a)} \qquad d\hat{z}_i/dD_i = 1/R^i_z(\hat{z}_i) > 0$$

$$\text{(3b)} \qquad d\hat{z}_i/dD_j = 0$$

$$\text{(3c)} \qquad d\hat{z}_i/dq_i = -R^i_i(\hat{z}_i)/R^i_z(\hat{z}_i)$$

$$\text{(3d)} \qquad d\hat{z}_i/dq_j = -R^i_j(\hat{z}_i)/R^i_z(\hat{z}_i) > 0.$$

The natural assumption, abstracting from agency problems between managers and shareholders, is that managers maximize equity value in this stage of the game when debt levels are taken as given. Later in the paper we examine the earlier decision of how much debt the firm should take on. At this earlier stage, managers are assumed to maximize total value. Once the debtholders are captive, however, the managers have no subsequent incentive to act in the debtholders interests. For purposes of comparison it is, nevertheless, useful to consider the problem of maximizing the debt value of the firm in the output phase, as if debtholders were running the firm. In this case, the maximand, denoted $W^i(q_i, q_j;.)$, is given by

$$\text{(4)} \quad W^i(q_i, q_j;.) = \int_{\underline{z}}^{\hat{z}_i} R^i(q_i, q_j, z_i) f(z_i)\,dz_i + D_i(1 - F(\hat{z}_i)).$$

(We are assuming that $\underline{z}$, the lowest possible value of z_i, generates positive operating profits for all relevant values of the choice variables. This is a convenient but inessential assumption.) The first term in (4) represents the operating profit of the firm in states of the world when this profit is insufficient to completely cover debt obligations. The second term represents those states of the world in which the creditors of the firm are paid in full.

The model presented here is the simplest model we could develop to explore possible connections between financial decisions and oligopolistic output markets. Generalizations to include more firms, other forms of market rivalry, and correlated random disturbances yield the same qualitative results.

II. Output Market Equilibrium

This section examines how the limited liability aspects of financial leverage affect the strategic output decisions by firms. Taking existing debt levels D_1 and D_2 as predetermined, the management of each firm chooses output to maximize either V or W, depending on whether it acts in the interest of shareholders or debtholders.

A. *Equity Value Maximization*

We take the case of equity value maximization as our standard case. Assuming an interior solution, the choice of output for firm i is obtained by setting the derivative of (1) with respect to q_i equal to zero:[9]

$$\text{(5)} \qquad V^i_i = \int_{\hat{z}_i}^{\bar{z}} R^i_i(q_i, q_j, z_i) f(z_i)\,dz_i.$$

The second-order condition is

$$\text{(6)} \qquad V^i_{ii} < 0.$$

The Nash output (or Cournot) equilibrium is obtained from the simultaneous solution of (5) for $i, j = 1,2$. In addition we also require that

$$\text{(7)} \qquad V^i_{ij} < 0,$$

$$\text{(8)} \qquad V^i_{ii}V^j_{jj} - V^i_{ij}V^j_{ji} > 0,$$

---

[9] Besides the expression in (5), the derivative of V^i with respect to q_i also includes another term, $-d\hat{z}_i/dq_i(R^i(\hat{z}_i) - D_i)$, which vanishes by (2).

which are standard conditions in Cournot-type models. Equation (8) is equivalent (given second-order conditions) to reaction functions being downward sloping. It tends to hold if expected marginal revenue declines when the output of the other firm rises. Equation (8), if it holds everywhere, implies uniqueness of the equilibrium and reaction function stability. It is well-known that even in the simplest Cournot models, conditions analogous to (7) and (8) can be violated by feasible demand and cost structures, and that is certainly true here. One case in which these conditions do hold, however, is if z_i is uniformly distributed, demand is linear, and marginal cost is constant.

While this paper focuses nearly exclusively on the Cournot duopoly market structure, the first central insight we offer applies quite generally. This central insight is that higher debt levels tend, in the standard case ($R^i_{iz} > 0$), to increase a firm's desired output. The intuition is as follows.

In bad states of the world, the firm's profits are insufficient to meet its debts. The equity holders' claims on the firm go to zero, and the debtholders become the residual claimants on the firm's profits. In other words, limited liability implies that debtholders become residual claimants in bad states: specifically, when $z_i \leq \hat{z}_i$. Equity holders are residual claimants in good states of the world, as illustrated in Figure 1.

An increase in debt causes $\hat{z}_i$ to rise, meaning that the range of states over which the firm becomes bankrupt is expanded. In particular, with $R^i_{iz} > 0$, it is states with low marginal returns to output that are moved from the region in which equity holders are residual claimants to the bankrupt region, where debtholders are the residual claimants. In other words, these low marginal profit states are no longer relevant to equity holders and equity holders would therefore want output to rise. Thus an increase in debt tends to make equilibrium output rise.

A football team that is behind late in the game will take chances that it would not normally take. The reasoning is that bad realizations are irrelevant, for the worst the team can do is lose, and it will do that anyway if it does not take chances. The more the team is behind, the more aggressive it will become. Our firm, representing equity holders, reacts in the same way. As debt rises, low marginal value states become irrelevant, for in those states the firm is turned over to the debtholders, and the equity holders get zero in any case. Since the firm restricts attention to higher marginal profit states, it adopts a more aggressive stance. (Note that the reasoning is precisely reversed if $R^i_{iz} < 0$.)

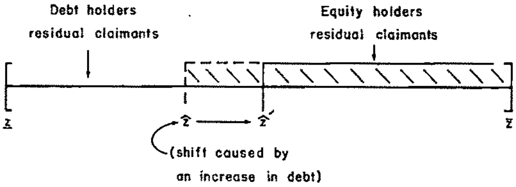

Debt holders residual claimants
Equity holders residual claimants
$\underline{z}$
$\hat{z}$
$\hat{z}'$
$\bar{z}$
(shift caused by an increase in debt)

FIGURE 1. DIVISION OF STATE SPACE INTO DEBT-RELEVANT AND EQUITY-RELEVANT REGIONS

Presenting this intuition rigorously requires a slightly different treatment for each different market structure. The case of monopoly is easiest. The monopoly case is obtained by letting $q_j = 0$ in expression (5) and by assuming that q_j remains at zero throughout the comparative static exercise. Total differentiation of (5) with respect to q_i and D_i then yields the comparative static formula:

$$dq/dD = -V^i_{iD}/V^i_{ii}.$$

The denominator is negative by second-order condition (6), which means that dq/dD has the same sign as V^i_{iD}: Output rises with debt if increases in debt cause marginal expected profits to rise.

The expression for V^i_{iD} is given by

$$V^i_{iD} = -R^i_i(\hat{z}_i)\,d\hat{z}_i/dD,$$

or, using (3a),

$$V^i_{iD} = -R^i_i(\hat{z}_i)/R^i_z(\hat{z}_i)f(\hat{z}_i). \tag{9}$$

The intuition associated with this expression is central to our paper. The denominator of (9) is obviously positive. Therefore, V^i_{iD} (and dq/dD) have the opposite sign to $R^i_i(\hat{z}_i)$.

This term is just marginal profit evaluated at the worst state of nature relevant to equity holders. With R_i^i increasing in z, it follows that R_i^i evaluated at $\hat{z}$ must be negative, since a weighted average of R_i^i over $\hat{z}$ and strictly better states is zero from first-order condition (5). Expression (9) implies that V_{iD}^i and, therefore, dq/dD must be positive in this case. The point is that $\hat{z}$ is precisely the marginal state that is being removed from consideration (by equity holders) as D rises. Consequently, the associated value of marginal profit, $R_i^i(\hat{z})$, is also thrown out of the marginal calculation determining output. Since this is the lowest value of marginal profit, optimal output must rise.

This discussion proves that output is rising in debt (for $R_{iz}^i > 0$) in the case of monopoly. This can also be proven for perfect competion and for other market structures as well. For reasons to be discussed later, it turns out that the cases of perfect competition and monopoly are not very interesting when the full two-stage game is considered. The duopoly case is interesting in the full game. We return to the duopoly case now. Occasionally in what follows it will be helpful to examine circumstances in which the firms are symmetric in the sense that $D_1 = D_2 = D$ and operating profits R^i and R^j are identical functions.[10] In this case, given (8), the Nash equilibrium in outputs will be symmetric and unique. Proposition 1 is a formal statement, for the symmetric duopoly case, of the result just described for monopoly.

PROPOSITION 1: *Assume firms* 1 *and* 2 *are symmetric. Then the Nash equilibrium output level* $q = q_i = q_j$ *is increasing in the debt level* $D = D_i = D_j$ *when* $R_{iz}^i > 0$ *and decreasing when* $R_{iz}^i < 0$.

For this and all subsequent proofs, we present the analysis only for the $R_{iz}^i > 0$ case, as the other case follows easily in a similar manner.

PROOF:

($R_{iz}^i > 0$) The comparative static effect of an increase in the common debt level, D, is determined by totally differentiating first-order condition (5) ($V_i^i = 0$) with respect to q_i, q_j, and D. Normally one would have to differentiate both first-order conditions (for i and j) and solve the resulting simultaneous system for comparative static effects dq_i/dD and dq_j/dD. Here, however, we can exploit the symmetry of the model and totally differentiate just one of the first-order conditions, incorporating the constraint that $dq_i = dq_j = dq$. This yields

$$V_{ii}^i dq + V_{ij}^i dq + V_{iD}^i dD = 0.$$

Solving for dq/dD then yields

$$(10) \qquad dq/dD = -V_{iD}^i/\left(V_{ii}^i + V_{ij}^i\right).$$

The denominator in (10) is negative by (6) and (7). Just as for the monopoly case, V_{iD}^i is given by

$$(11) \qquad V_{iD}^i = -R_i^i(\hat{z}_i)/R_z^i(\hat{z}_i)f(\hat{z}_i),$$

and, since R_i^i is increasing in z, $R_i^i(\hat{z}_i)$ must be negative (once again using (5)), implying that the expression in (11) is positive. Combining this with (10) yields

$$(12) \qquad dq/dD > 0,$$

as was to be shown.

Two important corollaries follow directly from Proposition 1.[11]

COROLLARY 1: *Assume firms* 1 *and* 2 *are symmetric. A completely equity-financed industry* ($D = 0$) *will produce a lower output than the corresponding leveraged industry*

---

[10] The symmetric case is also of empirical interest. Michael Bradley, Gregg Jarrell, and E. Han Kim (1984) present evidence that firms within the same industry tend to choose similar financial structures.

[11] The statement of Proposition 1, and Corollaries 1 and 2 are strictly correct provided $\hat{z}_i > \underline{z}$ whenever $D_i > 0$. For simplicity we assume this to be the case in all that follows. If $z_i \leq \underline{z}$ for some range of positive debt levels, then the output of firm i would remain unchanged over that range.

($D > 0$) when $R^i_{iz} > 0$. If $R^i_{iz} < 0$, the equity-financed industry will produce more than the leveraged industry.

COROLLARY 2: *A necessary and sufficient condition for financial structure to have no effect on the output market is $R^i_{iz} = 0$ for $i = 1,2$.*

The intuition underlying Proposition 1 and Corollary 1 is just as described in the monopoly case, and is not offset by the interaction between firms.

It has been noted by several authors, particularly Jensen-Meckling and Myers (1977), that if investments are chosen after financial structure is set (in contrast to the assumption of fixed capital maintained in our paper), then increasing the debt level should cause equity holders to undertake more risky investments, since they can declare bankruptcy in bad states of the world, while earning high rates of return in good states.[12] Our results concerning the output decision are conceptually similar. The normal case, $R^i_{iz} > 0$, corresponds to a situation in which increasing output increases the variance in the firm's profit stream because the marginal returns from extra production are positively correlated with the overall fortunes of the firm: the firm benefits most from increased production in good states of nature, and is harmed most by overproduction in bad states of nature. Thus increasing output is, in our model, analogous to a risky investment in the Myers framework in that it tends to be more attractive to shareholders when the firm is partially debt financed. When $R^i_{iz} < 0$, increasing output is risk reducing. A debt-financed firm will, in this case, reduce output, as it has less need to avoid risk because it can declare bankruptcy in bad states of the world.

Corollary 2 confirms the assertion set out in the introduction that one cannot legitimately treat the financial and real sides separately, except in special cases. Our next result indicates the strategic commitment aspects of financial decisions on the firms' behavior in the output market.

PROPOSITION 2: *Given $R^i_{iz} > 0$, a unilateral increase in firm i's debt, D_i, causes an increase in q_i and a decrease in q_j. If $R^i_{iz} < 0$, then $dq_i/dD_i < 0$ and $dq_j/dD_i > 0$.*

PROOF:

($R^i_{iz} > 0$) The method of proof is to totally differentiate first-order conditions (5) to generate the following system:

$$V^i_{ii}dq_i + V^i_{ij}dq_j + V^i_{iD_i}dD_i = 0, \tag{13}$$

$$V^j_{ji}dq_i + V^j_{jj}dq_j + V^j_{jD_i}dD_i = 0. \tag{14}$$

The first point to note is that V^j_j does not depend on D_i. Putting (13) and (14) in matrix form and using Cramer's rule to solve for comparative static effects dq_i/dD_i and dq_j/dD_i yields

$$dq_i/dD_i = -V^i_{iD_i}V^j_{jj}/B \tag{15}$$

$$dq_j/dD_i = V^i_{iD_i}V^j_{ji}/B, \tag{16}$$

where $B = V^i_{ii}V^j_{jj} - V^i_{ij}V^j_{ji} > 0$ from (8). Since B is positive, while $V^j_{jj} < 0$ by (6) and $V^j_{ji} < 0$ by (7), all that is needed is to sign $V^i_{iD_i}$.

$$\begin{aligned} V^i_{iD_i} &= -R^i_i(\hat{z}_i)f(\hat{z}_i)d\hat{z}_i/dD_i \\ &= -R^i_i(\hat{z}_i)/R^i_{z_i}f(\hat{z}_i), \end{aligned} \tag{17}$$

by (3a). Notice that (17) is identical to (11) which we have already established is strictly positive in the proof of Proposition 1. It follows, therefore, that

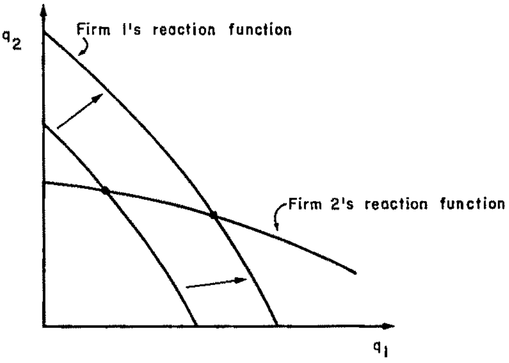
$$dq_i/dD_i > 0; \quad dq_j/dD_i < 0, \tag{18}$$

which completes the proof.

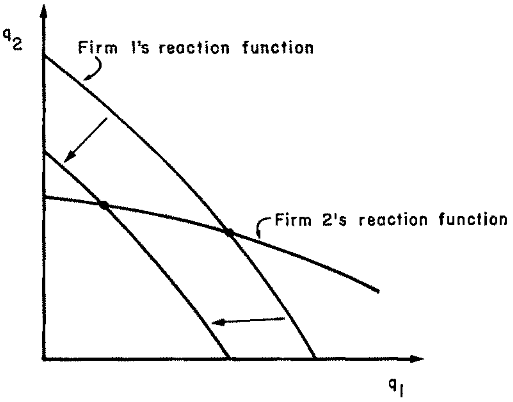
Proposition 2 represents the key insight to be brought out in this analysis. Notice that first-order condition (5) is the reaction function for firm i in implicit form indicating

---

[12] See also D. Galai and R. W. Masulus (1976) and Joseph Stiglitz and Andrew Weiss (1983) for related consideration of conflict of interest issues.



(a) $R^i_{iz} > 0$



(b) $R^i_{iz} < 0$

FIGURE 2

firm i's optimal output as a function of q_j, the output of its rival. The position of firm i's reaction function in output space depends on the debt level of firm i. In particular, with $R^i_{iz} > 0$, higher levels of debt, D_i, make it optimal for firm i to produce more in response to any output from its rival, firm j. As explained earlier in connection with Figure 1, an increase in a firm's own debt level removes states of low marginal return from the region in which equity holders are residual claimants. This translates into a higher desired output level for any given output level chosen by a rival: in other words, as illustrated in Figure 2a, the reaction function is shifted out. In effect, with $R^i_{iz} > 0$, debt financing serves to commit the firm to an aggressive stance in the output market.

B. *Debt Value Maximization*

The idea that managers of a firm might be controlled by debtholders cannot be taken as a serious representation of many North American firms, but it may have some empirical significance.[13] What we wish to do here is compare the objectives of debtholders with those of the equity holders.

Proceeding as before, we characterize the Nash equilibrium in the output market, where now firms act to maximize their debt value as given by (4). The Nash equilibrium output levels are given by the simultaneous solution to

$$(19)\quad W^i_i = \int_{\underline{z}}^{\hat{z}_i} R^i_i(q_i, q_j, z_i) f(z_i)\, dz_i = 0,$$

for $i = 1, 2$.[14] In addition to (19) we require

$$(20)\qquad W^i_{ii} < 0,$$

$$(21)\qquad W^i_{ij} < 0,$$

$$(22)\qquad W^i_{ii}W^j_{jj} - W^i_{ij}W^j_{ji} > 0.$$

These conditions carry the same interpretation as (6)–(8) for the case of equity maximization. According to Proposition 1, equity-managed firms tend to produce a level of output above the level that maximizes overall firm value (debt value plus equity value). As noted by a referee, it follows almost directly that a firm managed by debtholders would choose an output level below the output maximizing level. We state this result without proof[15] as Proposition 3.

PROPOSITION 3: *Assume firms* 1 *and* 2 *are symmetric and* $D_1 = D_2 > 0$. *Equilibrium output under equity value maximization is*

---

[13] For example, in some situations, control by debtholders, especially banks, might be a precondition for obtaining financing.

[14] Besides the term in (19), W^i_i also contains the term $(d\hat{z}_i/dq_i)(R^i(\hat{z}) - D_i)$, which is zero by (2).

[15] A rigorous proof of Proposition 3 is contained in an earlier version of our paper available as UBC Economics Discussion Paper 85–10, or from the authors.

greater than (less than) equilibrium output under debt value maximization when $R^i_{iz} > 0$ $(R^i_{iz} < 0)$.

The results of Proposition 3 indicate the conflict of interest that exists between bondholders and equity holders in the firm. The different preferences of debt and equity holders for output levels are easily explained by noting that equity holders are residual claimants in some states, while debtholders are residual claimants in other states, as illustrated in Figure 1. When $R^i_{iz} > 0$, equity holders prefer larger output levels than do debtholders, because the equity holders are the residual claimants in states of nature when marginal returns are high. Debtholders become the residual claimants in those states when the firm cannot cover its debt obligation, which happen to be states in which the marginal returns to extra output are very low.

Rather clearly, if $R^i_{iz} < 0$, so that high marginal returns to output are associated with bad states of the world in which the debtholders are residual claimants, the results is reversed. In such a case, debtholders would prefer higher output levels than equity holders.

An interesting implication of the foregoing discussion is that when $R^i_{iz} > 0$, equity holders in the industry might well be better off if firms in the industry were controlled by bondholders. The reason for this is that equity-managed firms will tend to produce more than the joint profit-maximizing output. If the firms could agree to act as a cartel, they could increase combined profits. Transferring ownership to debtholders would cause the firms to move toward the cartel output level. Of course, no single firm would by itself have an incentive to transfer control, for if one firm did and the other did not, the firm controlled by the debtholders would do even worse than at the original equilibrium. In effect equity-controlled leveraged firms suffer a magnified version of the usual "prisoner's dilemma" aspect of oligopolistic rivalry, with the amount of magnification increasing with leverage.

It is intuitively reasonable that in the standard case $(R^i_{iz} > 0)$, debtholders would prefer lower output levels than equity holders. What is perhaps more surprising is that debtholders, like equity holders, will in this case increase their desired output level as debt levels increase. In general comparative static effects for equity-controlled industries and debt-controlled industries are qualitatively identical.

PROPOSITION 4: *A debt-controlled industry has the following comparative static properties*: (a) *If firms are symmetric, equilibrium output is increasing in the common debt level* D *for* $R^i_{iz} > 0$, *and it is decreasing in* D *for* $R^i_{iz} < 0$. (b) *With* $R^i_{iz} > 0$, *a unilateral increase in the debt of firm* i, D_i, *causes an increase in the output of firm i and a decrease in the output of firm j. If* $R^i_{iz} < 0$, *the signs are reversed.*

PROOF:

Proofs are obtained by totally differentiating first-order conditions (19) with respect to output levels and debt levels and, solving for the comparative static effects, then using (20), (21), and (22), along with (3) to obtain signs. The details of the proofs are not reported here as they are virtually identical to the details in the proofs of Propositions 1 and 2.

Although these results may seem counterintuitive at first glance, the interpretations are straightforward. Consider the $R^i_{iz} > 0$ case. Looking at Figure 1 it is clear that an increase in debt levels raises the critical value of z_i at which bankruptcy occurs. Debtholders become the residual claimants over a wider range of states of the world, as higher levels of z are added to this range. Because $R^i_{iz} > 0$, the expected marginal profit of extra output goes up in the range relevant to debtholders. Therefore, on the margin it is profitable for the debtholders to increase output. As pointed out by a referee, in the limit, as the debt level becomes so large that debtholders become residual claimants in all states of the world, the firm will be managed just as if the firm were completely equity financed. Part b of Proposition 4 is explained by the same phenomenon that underlies the behavior of equity holders. For example, when $R^i_{iz} > 0$, a unilateral increase in leverage shifts the firm's reaction function outwards, increasing its equilibrium output and lowering the output of its rival.

III. Selection of Debt Levels

The preceding sections have examined the dependence of industry output levels on debt structure, treating debt as a predetermined or exogenous variable. In this section we describe the determinants of the debt structure. The existing literature on capital structure examines several important factors influencing the amount of debt chosen. The most standard treatment involves trading off the tax advantages of debt against bankruptcy costs in determining the optimal debt-equity position (as in Kraus and Litzenberger). Also, some analysts have stressed the use of capital structure to signal information about the firm to investors.[16] In this section we abstract from these well-understood determinants of financial structure and focus instead on an additional motive for holding (or not holding) debt that derives from the strategic commitment aspects of leverage in relation to output markets.

In what follows we assume that the manager of the firm is free to choose whatever output level he desires after debt is issued. In particular, bond convenants, which would restrict the manager's strategy decisions, are not considered. Bond convenants and other precommitment devices might, of course, be used for strategic purposes, but here we focus exclusively on strategic commitment through financial structure.

The equilibrium concept we use for the selection of debt levels is the Nash equilibrium in debt levels, subject to the constraint that firms and bondholders correctly anticipate the resolution of the Nash equilibrium in the output market. The equilibrium is, therefore, sequentially rational. More specifically, firms and investors understand that equilibrium output levels are determined by debt levels as implied by the simultaneous solution to first-order conditions (5). We denote the functional dependence of output levels on debt levels as follows: $q_i = q_i(\mathbf{D})$, where $\mathbf{D} = (D_1, D_2)$.

The objective of the firm's owners when making the debt decision is to maximize the total value of the firm, which is the sum of the equity value V^i and the debt value W^i. The basic point is that if potential debtholders are foresighted, then owners of the firm can sell bonds which promise to pay D_i only for their true value, taking into the account the possibility of bankruptcy. This true value is W^i. Hence the total value of the firm, denoted by Y^i is

$$(27) \quad Y^i\big(q_i(\mathbf{D}), q_j(\mathbf{D}), \mathbf{D}\big) = \int_{\underline{z}}^{\hat{z}_i} R^i\big(q_i(\mathbf{D}), q_j(\mathbf{D})\big) f(z_i)\,dz_i + \int_{\hat{z}_i}^{\bar{z}} R^i\big(q_i(\mathbf{D}), q_j(\mathbf{D})\big) f(z_i)\,dz_i.$$

This expression is obtained by adding (1) and (4) and noting that

$$\int_{\hat{z}_i}^{\bar{z}} D_i f(z_i)\,dz_i = D_i\big(1 - F(\hat{z}_i)\big),$$

leaving the two terms shown in (27). This combined value is, as one would expect, the expected value of operating profits over all states of the world. Because we abstract from bankruptcy costs and from the tax advantages of debt, issuing debt is strictly a break-even transaction for the firm, except for the fact that equilibrium output levels will depend on debt levels. If, by way of contrast, planned output levels were exogenously fixed (by, for example, quota allocations) then issuing debt would be purely neutral, having no effect on total value, Y^i, as in a Modigliani-Miller world. However, q_i and q_j are written as functions of $\mathbf{D}$, the industry debt structure, to indicate that output levels are in fact functions of the debt levels. A particular debt structure for the industry is a commitment to a particular output structure.

---

[16] Standard references on the use of capital structure to convey information include Stephen Ross (1977), Hayne Leland and D. Pyle (1977), and Robert Heinkel (1982).

The marginal effect of an increase in D_i on the value of firm i is given by

$$(28) \quad Y^i_{Di} = \left[\int_{\hat{z}}^{\bar{z}} R^i_i(z_i) f(z_i)\,dz_i\right] dq_i/dD_i + \left[\int_{\underline{z}}^{\hat{z}} R^i_i(z_i) f(z_i)\,dz_i\right] dq_i/dD_i + \left[\int_{\underline{z}}^{\hat{z}} R^i_j(z_i) f(z_i)\,dz_i + \int_{\hat{z}}^{\bar{z}} R^i_j(z_i) f(z_i)\,dz_i\right] dq_j/dD_i.$$

The first term is zero by (5). The second term measures the effect of an induced change on q_i on the debt value of the firm. Notice that when $R^i_{iz} > 0$, (5) implies that $R^i_i(z_i) < 0$ for all $z_i \leq \hat{z}_i$, reflecting the fact that equity holders will choose higher output levels than debtholders would like. A similar argument serves to establish that $R^i_i > 0$ for all $z \leq \hat{z}$ whenever $R^i_{iz} < 0$. Taken together these two conditions imply

$$(29) \quad \int_{\underline{z}}^{\hat{z}} R^i_i(z_i) f(z_i)\,dz_i < 0 \,(>0) \quad \text{if } R^i_{iz} > 0 \,(<0).$$

Using Proposition 2 ($dq_i/dD_i > (<0)$ if $R^i_{iz} > (<) \; 0$), combined with (29), allows us to sign the second term in (28):

$$(30) \quad \left[\int_{\underline{z}}^{\hat{z}} R^i_i(z_i) f(z_i)\right] dq_i/dD_i < 0.$$

This term indicates that the induced change in output caused by taking on more debt exacerbates the conflict of interest between debt and equity holders and lowers the debt value of the firm. The third and final term of expression (28) represents the strategic effect of debt. A higher debt level for firm i induces a change in the equilibrium output of firm j. Specifically, if $R^i_{iz} > 0$, a higher level of debt for firm i implies lower output for firm j. This effect, taken by itself, raises both the debt value and the equity value of firm i because $R^i_j < 0$. A lowered output by firm j is unambiguously good for firm i.

Thus there are two conflicting effects of increasing debt on the value of the firm. Extra debt worsens the conflict of interest between debt and equity holders, tending to lower the value of the firm, as reflected in (30). In addition, however, extra debt has the value-increasing strategic effect we have just described. Despite these partially offsetting effects of debt, we are able to prove that the strategic effect dominates for sufficiently small levels of debt, insuring an interior ($D^i > 0$) solution to the firm's value maximization problem. Specifically, if $D_i = 0$ there is no conflict between bondholders and equity holders (there are no bondholders) implying that the second term of (28) is zero.[17] The first term is of course equal to zero, while the third term remains strictly positive implying that $Y^i_{D_i}$ is strictly positive at $D_i = 0$.

The other case, $R^i_{iz} > 0$, has strikingly different implications. In this case, dq_j/dD_i is positive, which implies that increases in D_i will cause the rival firm to increase output, which in turn lowers profits for firm i. Therefore the third term of (28) is strictly negative. Expression (28) as a whole is strictly negative at all feasible (nonnegative) levels for debt and we obtain a corner solution at $D^i = 0$. These results are expressed in Proposition 5.

PROPOSITION 5: *Industry debt levels will be strictly positive if* $R^i_{iz} > 0$. *If* $R^i_{iz} < 0$, *firms will be entirely equity financed.*

Proposition 5, together with Proposition 1, implies that if firms hold any debt, they will produce more output than in the traditional industrial organization version of oligopoly, in which firms are assumed to be 100 percent equity financed.

The intuition underlying Proposition 5 is derived from our earlier results in Propositions 2 and 3. When marginal returns from output are positively correlated with the state of the world ($R^i_{iz} > 0$), Proposition 2 indicates that taking on more debt confers a strategic advantage of the firm, in that it causes the rival's equilibrium output to fall. This strategic benefit from debt financing must be traded off against the resulting decrease in the debt value of the firm as debt increases (as indicated by expression (31)). We have shown that the strategic effect must dominate for sufficiently low debt levels, leading to an internal solution for the debt-equity ratio. In the case in which good states of the world are correlated with low marginal returns to output, debt has only value-reducing effects. The strategic effect of debt causes rival output to rise, and the conflict of interest between debtholders and equity holders remains. A corner solution with no debt is the solution in this case.

---

[17] Strictly speaking $\hat{z} = \underline{z}$ in this situation, rendering the integral equal to zero.

It also follows from the discussion above that a monopolist or a perfectly competitive firm would choose a corner solution with no debt. In either case, the strategic effect of debt financing is not relevant and firms would have no reason, in our model, to use debt. Therefore, strategic commitment through financial structure does not apply to monopoly and perfect competition in the same way that it applies to oligopoly.

Proposition 5 should be interpreted with some care, since it is derived from a model in which certain empirically important determinants of financial structure, such as taxes, are explicitly ignored. Our analysis isolates the strategic output market effects as an influence on financial structure.

Our final result makes the point that, at least in the symmetric case, equilibrium debt levels do not maximize value for the firms taken together. While this is not surprising, given that firms behave noncooperatively in financial markets, it does suggest certain collusive financing arrangements that firms might profitably engage in.

PROPOSITION 6: *Under symmetric conditions the value of the industry,* $Y^1(\mathbf{D}) + Y^2(\mathbf{D}) = 2Y(\mathbf{D})$, *is not maximized in equilibrium. In particular,* $dY(\mathbf{D})/dD < 0$ *if* $R^i_{iz} > 0$, *and* $dY(\mathbf{D})/dD > 0$ *if* $R^i_{iz} < 0$.

PROOF:

($R^i_{iz} > 0$) If (6) and (8) hold globally, then in the special case where $D_i = 0$, $\hat{z} = \bar{z}$, and $q_i = q_j = q$, then (6) and (8) imply

$$\int_{\underline{z}}^{\bar{z}} R^i(q, q, z_i) f(z_i)\, dz_i \tag{33}$$

is strictly concave in q. Proposition 6 then follows directly from the fact that completely equity-financed firms fail to maximize joint profits because of excess production. This tendency on the part of firms to overproduce is exacerbated as they take on debt.

The basic point of Proposition 6 is straightforward. A noncooperative oligopoly produces more output than a profit-maximizing cartel or monopoly would. When $R^i_{iz} > 0$, increases in debt beyond zero cause output to rise still further: the use of debt lowers profits. In other words, debt is actually procompetitive. An interesting observation is that standard Cournot oligopoly corresponds exactly to the case of complete equity financing. Ignoring the interaction between financial and output markets causes the competitiveness of such oligopolies to be understated.

This structure also suggests that central control of financing arrangements might be an attractive collusive practice. If credit markets for a particular industry are quite concentrated, then lenders would have incentives to act as facilitating agents for collusion.

IV. Concluding Remarks

This paper makes the basic point that product market decisions and financial decisions will normally be related. We have analyzed this relationship for a particular industry structure in which financial decisions and product market decisions follow in sequence. In this situation, the limited liability provisions of debt financing imply that changes in financial structure alter the distribution of returns between debt and equity holders, and therefore change the output strategy favored by equity holders.

Because financial structure influences the output market equilibrium, foresighted owners of the firms will have incentives to use financial structure precisely so as to influence the output market in their favor. Given the behavior of the rival firm, a firm which ignored the strategic effect of financial decisions would have lower total value than a firm which took advantage of these effects.[18] These strategic uses of financial structure are purely predatory, and the net effect when both firms use them is that both firms are worse off. In our model, symmetric firms use financial structure as a commitment variable to influence the output market equilibrium. The symmetry is, however, not fundamental. One could examine an asymmetric market in which, for example, an incumbent firm used financial structure to preempt possible entry by a rival, just as capital or *R&D* can be used to deter entry.

Our analysis abstracts from empirically important aspects of the financial structure decision. In any empirical work one would certainly have to incorporate the tax advantages of debt, and the possibility of bankruptcy costs. In addition, the so-called agency aspects of financial structure could be very important. Specifically, if the firm is run not by shareholders but by imperfectly monitored managers, then very different results could emerge. Managers, presumably, would be very concerned about losing their jobs, especially if the outside world cannot tell whether a bankrupt firm has suffered bad luck or bad management. If so, then high debt levels might normally make the management of a firm extremely cautious and might tend to reduce industry output, in contrast to certain results in our model.

The analysis by Michael Bradley et al. and others indicates that there are systematic differences across industries and similarities within industries with respect to financial structure. Our analysis suggests that these variations in financial structure might be explained by industry-specific factors. For example, the mode of competition within an industry: price competition, quantity competition, *R&D* races, competitive advertising, and so on, would all have significant and different implications for financial structure. In addition, as shown in our model, the pattern of random returns (as reflected, for example, by the sign of R^i_{iz}) is likely to have an effect on financial structure.

Another implication of our analysis is that the institutional structure of credit markets can have an important impact on the economic performance of output markets. In our model, we identify an opportunity for credit institutions with monopoly power to act as a facilitating agent for collusion in the output market, but this is only one of several possible links. Finally, our model suggests that the public finance aspects of interest deductability might well include the output effect that is induced by the resulting higher debt levels.

The overall point that we wish to emphasize is that opening the linkage between financial markets and the "real" side of firms' decisions for analysis suggests a number of possibly important and certainly interesting economic consequences, of which our model provides some examples.

REFERENCES

Aivazian, Varouj A. and Callen, Jeffrey L., "Corporate Leverage and Growth: The Game Theoretic Issues," *Journal of Financial Economics*, December 1980, *8*, 379–99.

Allen, Franklin, "Capital Structure and Imperfect Competition in Product Markets," Discussion Paper, University of Pennsylvania, 1985.

Altman, Edward I., "An Empirical Investigation of the Bankruptcy Cost Question," *Journal of Finance*, September 1984, *39*, 1067–89.

Bradley, Michael, Jarrell, Gregg A. and Kim, E. Han, "On the Existence of an Optimal

---

[18] The difference between our analysis and the analysis of conflict of interest in Myers (1977) is that here the returns themselves are endogenous. The endogeneity of the returns arises because the returns depend on a choice variable, output. If returns were purely exogenous, as in Myers, the creation of a conflict of interest between bondholders and debtholders could only lower the value of the firm unless there were some other advantage to debt, such as a tax advantage.

Capital Structure: Theory and Evidence," *Journal of Finance*, July 1984, *39*, 857–80.

Brander, James A. and Lewis, Tracy R., "Bankruptcy Costs and the Theory of Oligopoly," mimeo., University of British Columbia, 1985.

______ **and Spencer, Barbara J.,** "Strategic Commitment with R & D: The Symmetric Case," *Bell Journal of Economics*, Spring 1983, *14*, 225–335.

Bulow, Jeremy and Shoven, John, "The Bankruptcy Decision," *Bell Journal of Economics*, Autumn 1978, *9*, 437–56.

Dixit, Avinash, "The Role of Investment in Entry Deterrence," *Economic Journal*, March 1980, *90*, 95–106.

Eaton, B. Curtis and Eswaren, Mukesh, "Asymmetric Equilibria and Strategic Behaviour in Oligopolistic Markets: A Synthesis," Discussion Paper No. 84–19, University of British Columbia, 1984.

______ **and Lipsey, Richard,** "The Durability of Capital as a Barrier to Entry," *Bell Journal of Economic*, Autumn 1980, *11*, 721–29.

Galai, D. and Masulis, R. W., "The Option Pricing Model and the Risk Factor of Stock," *Journal of Financial Economics*, January/March 1976, *3*, 53–82.

Giammarino, Ronald M., "On the Significance of Financial Distress to the Theory of Optimal Capital Structure," unpublished doctoral dissertation, Queen's University, 1984.

Green, Jerry and Shoven, John, "The Effects of Financing Opportunities and Bankruptcy on Entrepreneurial Risk Bearing," in J. Ronen, ed., *Entrepreneurship*, Lexington: Lexington Books, 1983.

Heinkel, Robert, "A Theory of Capital Structure Relevance Under Imperfect Information," *Journal of Finance*, December 1982, *37*, 1141–50.

Jensen, M. and Meckling, W., "Theory of the Firm: Managerial Behavior, Agency Costs, and Capital Structure," *Journal of Financial Economics*, October 1976, *3*, 305–60.

Kraus, Alan and Litzenberger, Robert H., "A State-Preference Model of Optimal Financial Leverage," *Journal of Finance*, September 1973, *33*, 911–22.

Leland, Hayne and Pyle, D., "Information Asymmetries, Financial Structure and Financial Intermediation," *Journal of Finance*, May 1977, *32*, 371–87.

Maksimovic, Vojislav, "Optimal Capital Structure in Oligopolies," unpublished doctoral dissertation, Harvard University, 1986.

Modigliani, Franco and Miller, Merton, "The Cost of Capital, Corporation Finance, and the Theory of Investment," *American Economic Review*, June 1958, *48*, 261–97.

______ **and** ______, "Corporate Income Taxes and the Cost of Capital: A Correction," *American Economic Review*, June 1963, *53*, 433–43.

Myers, Stewart C., "Determination of Corporate Borrowing," *Journal of Financial Economics*, December 1977, *4*, 147–75.

______, "The Capital Structure Puzzle," *Journal of Finance*, July 1984, *39*, 575–92.

Ross, Stephen, "The Determinants of Financial Structure: The Incentive-Signalling Approach," *Bell Journal of Economics*, Spring 1977, *8*, 23–40.

Schelling, Thomas, "An Essay on Bargaining," *American Economic Review*, June 1956, *46*, 557–83.

Schmalensee, Richard, "A Model of Promotional Competition in Oligopoly," *Review of Economic Studies*, November 1976, *43*, 493–507.

Selten, R., "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 1975, *4*, 25–55.

Stiglitz, Joseph and Weiss, Andrew, "Credit Rationing in Markets with Imperfect Information," *American Economic Review*, June 1983, *73*, 393–410.

Titman, Sheridan, "The Effect of Capital Structure on a Firm's Liquidation Decision," *Journal of Financial Economics*, March 1984, *13*, 137–51.

Horizontal vs. Vertical Information Structure of the Firm

By MASAHIKO AOKI*

This paper compares the efficiency of two information structures, hierarchical and vertical, of the firm in coordinating operational decisions among technologically interrelated shops whose costs are uncertain. In the first, the capability of management to monitor and respond to emergent events at the shop level is bounded. In the second, production decisions are coordinated among shops without the centralization of information, whose capability of semiautonomous problem solving improves through learning by doing. This study is motivated by a U.S.-Japan comparison of industrial organization.

The objective of this paper is to compare the efficiency of two informational structures of the firm in coordinating operational decisions among interrelated units (shops) whose cost conditions are uncertain. One informational structure is hierarchical—management possesses a perfect a priori knowledge of the technological possibilities of shops, but is incapable of perfect monitoring of emerging events affecting these technologies, and/or having rapid corrective actions implemented at shops. The second structure is horizontal—production decisions are coordinated among semiautonomous shops that have only incomplete knowledge of technology at the outset, but gradually become capable of responding to emerging events more quickly by better uses of on-the-spot knowledge. This paper is exclusively concerned with efficiency properties of the two informational structures; and the issue of their incentive compatibilities is relegated to other work.[1] The comparison of the two structures is purely analytical, but I hope to provide a conceptual framework for a U.S.-Japan comparison of industrial organization (such as workshop organization, internal firm organization, manufacturer-suppliers relations, etc.). Another obvious application would be to production coordination in the socialist economy. In this application, the management should be read as the ministry of industry and shops as firms.

I. Motivation of the Study

This section, as a way of motivating the present study, describes several stylized contrasts between Japanese and U.S. firms pertinent to the operation of the internal organization, as well as to the degree and mode of integration. These references are mainly directed toward large unionized firms in the manufacturing industry.

A. *Specialization vs. Learning*

There is a distinct difference in the way work is organized on the shop floor in the typical Japanese firm (denoted J firm) and the typical American (unionized) firm (denoted A firm). In the A firm, workers' jobs are specified according to a job classification scheme stipulated in a collective agreement,

---

*Professor of Economics, Stanford University, Stanford, CA 94305 and Kyoto University, Kyoto, Japan, 606. I thank Kazuo Koike for discussions which inspired the present study. I am indebted to readers of an earlier version for helpful comments: Kenneth Arrow, James Hamilton, Hideshi Ito, Burton Klein, Hajime Miyazaki, David Mowrey, Michael Riordan, and Oliver Williamson; to referees of this *Review*, and to participants of seminars at Stanford, UC-San Diego, British Columbia, Washington, Tokyo, Ohio State, and the Stockholm School of Economics. Research leading to this manuscript was partially supported by grants from the Center for Economic Policy Research and the Center for Research in International Studies at Stanford University.

[1] See my 1985 paper for an exploratory attempt on this issue.

and each worker is required to perform his (her) own specialized task according to formal or informal rules, operation manuals, supervisor's directives, etc. When irregular events happen, such as the absenteeism of workers, breakdown of machinery, or production of an abnormally high rate of defective products, remedies are usually sought by supervisors, reliefmen, repairmen, engineers, and the like. Manual workers are not specifically responsible for coping with unexpected emergent events.

In contrast, in the *J* firm, workers' jobs are not specified in detail and workers rotate among various jobs with some frequency within, as well as beyond, workshops. Through this practice, workers are gradually made familiar with the whole work process and become capable of coping with unexpected emergencies. As Kazuo Koike convincingly argues, such capability constitutes one of the most important components of the workers' skills.[2] In fact, there is a considerable degree of delegation of decision-making power to the group of workers on the shop floor. They are encouraged to solve emergent problems by themselves and improvise improvements on designed work processes. When needed, the repairmen, inspectors, and engineers are willing to cooperate with the workers in joint problem solving.

The emphasis at the *A* firm is on the efficiency attained through job specialization and rational hierarchical control, whereas at the *J* firm, it is on the workers' grass roots capability to cope with emergent events fostered by collective learning by doing. The Japanese approach has proved to be effective from a quality- and cost-control point of view in industries such as steel and automotive manufacturing where cost reduction has been attained largely through improvements on fairly established technology. But, will it be equally effective in the high-tech industry where the speed of technological obsolescence is likely to be high, and general educational training seems to be gaining importance as a component of workers' skills?

B. *Hierarchical Control vs. Horizontal Coordination*

In the uncertain world where complete contingent contracts cannot be written because of "bounded rationality" of concerned agents, emergent events have often to be dealt with, *ex post*. Different modes of transactions would emerge, depending upon how and by whom they are dealt with. In the *A* firm, decisions involving discretion are normally hierarchically organized according to their importance. In fact, some authors argue that the layering of specialized decision making in order to cope with emergent events for which detailed specification of appropriate actions cannot be formulated *ex ante* is the essence of hierarchy.[3]

In the *J* firm as well, decisions involving a high degree of uncertainty, such as investment and research and development, as well as those responding to a high degree of irregularity are placed under hierarchical control. However, once an overall framework for production is laid down by such strategic decisions, horizontal informational exchanges and semiautonomous coordination of operations by relevant subordinates are emphasized. In contrast to the prominent role played by the expediter in the *A* firm, the coordination of production between *J* firm workshops to facilitate a smooth production flow along the production stream is often done by horizontal communication without the intervention of a supervisor, as the highly publicized "kanban-system" at the Toyota factory exemplifies. A kanban is a tag-like card put in a vinyl envelope which orders from the immediately upstream shop (or the supplier) a particular amount of materials, tools, parts, or processed goods at a particular time. Robert Cole aptly draws an analogy between this practice and "the system we use in restocking supermarkets" (1985, p. 106). Although apparently crude, the kanban system is considered to have contributed greatly to reducing inventory and waste, uncovering bottlenecks, allowing for rapid adjustment of product lines, and reducing managerial personnel.[4]

---

[2] The practice of job rotation in the Japanese workshop is described and its implication to skill formation is thoroughly discussed in Koike (1984). Based upon his own extensive field work, Koike (1985) also maintains that an important part of the workers' skills consists of their ability to cope with unexpected emergencies.

[3] See for example, David Kreps (1984).

The emphasis at the *A* firm is on the efficiency attained by rational technocratic control, whereas, in the *J* firm, it is the efficiency attained through the use of on-the-spot knowledge and rapid problem solving through learning by doing. In the former, decisions may be made consistently and rationally from the organizational point of view, once an emergent problem is recognized. However, if the quality of information is poor, the manager may be slow and imprecise in recognizing the problems to be handled. Also, subordinates, by not being included in decision making, may lack the motivation to report problems and to implement the hierarchical order in a precise and swift way. Thus using the hierarchical system involves the cost of monitoring due to the bounded rationality of the supervisors, and the cost of implementation due to the lack of incentives of the subordinates.

On the other hand, in the *J* firm, subordinates may be motivated to respond to emergent events swiftly in a way which they conceive to be consistent with the organizational purpose, provided that organizational goal is internalized by them. However, the ability of subunits to coordinate their decisions between themselves in a way actually consistent with the organizational purpose would be limited by their partial understanding of the whole mechanism operating within the firm. This understanding can be enhanced by learning by doing, but is costly in terms of time. Also, the problem-solving capability of subunits through horizontal coordination would be limited by the lack of centralization of information concerning emergent events which affect various other subunits.

Which of the two information and decision-making structures, hierarchical or horizontal, is more effective for controlling the total costs of the system in changing environments? How does the complexity of technological interdependencies of subunits affect their relative efficiency performances?

C. *Integration vs. Quasi Disintegration*

The degree of integration of the *A* firm is normally much higher than that of the *J* firm. A joint U.S.–Japan study of the automotive industry estimates that 45 percent of a car's purchased value is provided by U.S. manufacturers and their wholly owned subsidiaries, with 55 percent being provided by outside suppliers. Only 25 percent of a car's purchase value is made in-house for Japanese manufacturers.[5] This difference is due to an extensive reliance of Japanese major manufacturers on suppliers which are normally smaller in size and often referred to as "subcontractors."[6]

A reason often cited for the extensive deployment of the subcontracting relations by the *J* firm is that, since the rate of its growth was so fast, it could not generate enough internal resources for integration to an extent comparable to that which the *A* firm had achieved over its longer course of growth. This may partially explain the genesis of the extensive subcontracting relations observed in the Japanese manufacturing industry. But if this was the only reason, and there was no intrinsic comparative advantage of a subcontracting relationship over more complete integration, the former would have been gradually replaced by the latter as the growth rate of the economy slowed down and the accumulation of internal resources of the *J* firm progressed in the 1970's and the early 1980's. On the contrary, the degree of reliance of the *J* firm on subcontracting relationships has increased over this period.[7] Furthermore, the relationship between the parent firm and its subcontractors has come to be much more systematic and formalized. The large *J* firm maintains long-term, direct relations with primary subcontractors that are loosely organized into associations of cooperating firms (kyoryoku-kai); and those primary subcontractors in turn maintain relations with secondary subcontractors, etc. Instead of a highly integrated, hierarchical firm, we observe a quasi-disintegrated, quasi-hierarchical group of firms (hereafter referred to as the *J*-firm group) in which subcontractors maintain semiautonomous control of production, although the main manufacturers typically have minority equity holdings in their first-tier suppliers, and exercise substantial influence over the direction of research and development as well as over investment decisions by the latter.

---

[4] For more details on the kaban system, see Cole (1985) and Yashuhiro Monden (1983).

[5] See Cole and Taizo Yakushiji (1984, p. 153–54).

[6] Precisely speaking, the subcontractor is a legal concept in Japan. According to Basic Law for Medium and Small Enterprises and Law for Preventing the Delay of Payment of Subcontracting Fees, firms with 300 or fewer employees, or with 100 million yen or less paid-in capital, are classified as subcontactors when they have contractual relations with larger firms for supplying parts, processed products, matrials, etc. Excellent studies on subcontracting relations in the Japanese automotive industry are found in Banri Asanuma (1985) and Cole and Yakushiji (ch. 9). Also see my book (1984a).

The benefits of integration vis-à-vis the market have been discussed by many authors. Integration may be able to reduce the cost of market transactions caused by the opportunistic and inefficient behavior of a seller and buyer in situations where either the buyer or the seller must make transaction-specific investments (Oliver Williamson, 1975, and Benjamin Klein et al., 1978), and/or where it is too costly for one party to specify a long list of the particular rights it desires over another party's assets (Sanford Grossman and Oliver Hart, forthcoming). Another motive for integration may be to communicate uncertain information regarding upstream supply conditions more efficiently (Kenneth Arrow, 1976).

But as Grossman and Hart argue, the transaction-cost and information-cost-based arguments did not elucidate where the benefits of integration will stop. Ronald Coase (1937) states that the size of the firm is limited by the capacity of the single owner in the number of activities he can manage. This is unconvincing, however, because organizational innovation such as the multidivisional form may overcome this limit (Williamson, 1975). Facing this problem, Williamson (1985) came to argue that excessive integration may sacrifice the efficiency possible under the "high-powered" market incentive. Consequently, he states, if benefits of high-powered incentives are potentially high, while the transaction in question involves a substantial degree of asset specificity, "new hybrid forms" of organization other than the firm may appear in response. I submit that the information costs may also increase from excessive integration and that may provide another reason for the emergence of the *J*-firm group, which is an example of such a hybrid form.

The *J*-firm group may combine some of the benefits of both integration and the market. Since the relationship between the parent firm and subcontractors is of a long-term nature, both would come to have a fairly good knowledge of each other's technology, and inefficient price haggling would not be a serious problem. In Williamson's terminology, conditions of "information impactedness" (1975, pp. 31–33) are more easily overcome and, even when they appear, are less likely to give rise to strategic behavior. Also, since an implicit/explicit arrangement of profit sharing is the normal practice in this long-term relationship,[8] underinvestment in transaction-specific assets may be avoided to both parties' mutual advantage. On the other hand, since the long-term relationship is maintainable only if it is beneficial to both parties, but is otherwise dissolvable albeit at some cost, the "midpower" market incentive would operate within the *J*-firm group; that is, subcontractors would take greater care in cost control and quality control of their production as compared to in-house supply divisions. But these relative advantages still remain hypothetical, unless an efficient information system between the parent firm and its subcontrators for coordinating their operations is possible without consistent hierarchical direction under an integrated firm or the efficient signal of competitive prices.

---

[7] See my paper (1984b).

[8] See my paper (1985) and also Asanuma.

As indicated above, one of the premises for the efficiency of the *J*-firm group is technological knowledge sharing between the parent firm and its subcontractors for mitigating the problem of opportunistic behavior. However, within a general framework of knowledge sharing, day-to-day monitoring of emergent events affecting the cost efficiency of productive operations and/or product quality may be delegated to subcontractors. Then the information exchange necessary for rapid corrective action can be made more direct by coordination between the relevant operational units (for example, between parts-supplying subcontractors and the assembly shop of the parent firm, between first-tier and second-tier subcontractors, etc.), without the intermediary of an elaborate, centralized control mechanism

However, is this quasi-decentralized coordination within the *J*-firm group not a crude information system? Is it not likely to be the case that the administrative costs of a highly integrated firm will be reduced in the near future by the rapidly growing information-processing technology and more reliable robotic technology? Or will the quasi-disintegrated group continue to exhibit higher transaction-cost savings in spite of these new developments?

I have described three stylized contrasts between the *A* firm and the *J* firm/*J*-firm group. They include oversimplifications inevitable in any stylized formalization. Neither a pure *J* firm nor a pure *A* firm exists. Any American or Japanese firm can contain some aspects of both. There is also much anecdotal evidence that many American firms have recently been adopting some elements of the *J* firm, such as the simplification of the job classifications scheme, an increased delegation of decision-making authority to line organizations, more systematic relationships with suppliers, etc. One may also argue that the three issues discussed above are only mildly related to each other. I submit, however, that if these three stylized contrasting features are taken together as a benchmark from the informational structure point of view, one fundamentally common issue arises. That is the issue of relative efficiency of what I call "horizontal coordination" based on learning by doing vs. rational "hierarchical control" based on specialism.

Suppose that strategic decisions (such as those that involve investment as well as the direction of research and development) are made by superordinates (the management of the firm). Given these decisions, actual operations of subunits (shops) must be adjusted and coordinated accordingly as emergent events affect the efficiency of each unit. In order for efficient coordination to occur, the monitoring and identification of emergent events affecting cost efficiency as well as quick implementation of corresponding operational decisions are imperative. In the horizontal coordinating system, these tasks are delegated to subunits performing relevant operations that are limited in their capacity to recognize a relevant emergency at the outset. Even if they do, since the centralized use of information concerning emergent events is absent, problem solving at the subunit level may not satisfy the first-best condition. But, through learning by doing, the subunits can improve their ability to perceive emergent problems and to find corresponding solutions. Once a problem is recognized and a corresponding solution found, the subunits should be able to implement it quickly.

On the other hand, in the hierarchical control system, management has a priori knowledge of the overall production technology. But their capacity to identify emergent events affecting production technology of subunits and to enforce the implementation of appropriate operational decisions upon said subunits may be limited and costly, simply because management is removed from operational activities. In other words, the rationality of the superordinates is apt to be "bounded."

Thus the relative efficiency of the two systems depends upon the following factors: the ability of subunits to learn how technology is affected by emergent events and how these events should be dealt with (the speed of learning); the initially endowed ability of subunits to understand the nature of technology as well as the speed of obsolesence of technology; and the ability of management to identify emergent problems at the subunit level as well as to enforce the implementation of operational decisions upon subunits as quickly as possible (the quality of hierarchical informational structure and the reward system). In the next section, I construct a simple analytical model of horizontal coordination in which learning by doing by subunits would play a significant role, and a model of hierarchical control whose rationality is bounded; I then analyze how the factors affect the relative efficiency of the two models. As I have indicated, it is likely that the aspects of hierarchical control and those of horizontal coordination may coexist in any single firm. Therefore, the following model comparison may be understood as an analytical device for considering the problem of an optimal mix of the two approaches within a firm.

II. The Model

Let us employ the production model due to Jacque Cramer (1980) and John Geanakopolos and Paul Milgrom (1985) as a basis for the comparative analysis of the two informational structures of the firm. Let us consider a system composed of n technologically interrelated subunits called shops, identified by index $i=1,\dots,n$. Let $x_i \in R^s$ denote a vector of net production assignment of the ith shop, with the convention that negative components indicate inputs and positive components outputs. Suppose that the cost incurred by shop i to realize a production assignment x_i is represented by a quadratic cost function:[9]

$$C_i(x_i,u_i)=[x_i-u_i]'B_i[x_i-u_i]+A_i \qquad (i=1,\dots,n)$$

where B_i is a positive semidefinite matrix with at least one diagonal element being strictly positive, A_i is a constant matrix, and $u_i \in R^s$ is a random variable representing cost uncertainty with $E[u_i]=u_i^*$, $E[(u_i-u_i^*)(u_i-u_i^*)']=\operatorname{var} u_i$, and $E[(u_i-u_i^*)(u_j-u_j^*)']=0$ for $i \neq j$. The cost C_i may represent the "outlay" from the system or the effort expenditure level incurred by shop i to realize its production assignment.

The objective of the system is to realize a net output target:

$$x^*=\Sigma x_i$$

with the expected minimum aggregate cost

$$E[C(x,u)]=\Sigma E[C_i(x_i,u_i)]$$

responding to emergent events *ex post*. Given an observed event $u=(u_1,\dots,u_n)$, the solution to this problem is

$$x_i^p=u_i+B_i^{-1}B[x^*-U] \qquad (i=1,\dots,n) \tag{1}$$

where $B=[\Sigma B_i^{-1}]^{-1}$; $U=\Sigma u_i$.

If the system can identify emergent events and implement corresponding cost-minimizing solutions immediately, the expected total cost under this perfect control is

$$C^p=E\{\Sigma[x^*-U]'BB_i^{-1}B[x^*-U]\}+\Sigma A_i$$
$$=C^*+\operatorname{tr}[B\operatorname{var}U]$$

where $C^*=[x^*-U^*]'B[x^*-U^*]+\Sigma A_i$

$U^*=\Sigma u_i^*$

$$\operatorname{tr}[B\operatorname{var}U]=E\{[U-U^*]'B[U-U^*]\}.$$

On the other hand, if the system cannot identify emergent events, the system routinizes its production at the level

$$x_i^n=u_i^*+B_i^{-1}B[x^*-U^*] \qquad (i=1,\dots,n)$$

so as to realize the minimum expected total cost under no *ex post* information condition:

---

[9] Arrow (1985) cautions against the use of a quadratic function for the study of information structure. As he states, the use of the quadratic function sometimes leads to strange analytical results. I assume that the quadratic representation of technology is a local approximation for nonlinear technology.

$$C^n = \Sigma E\{[x_i^n - u_i]'B_i[x_i^n - u_i]\} + \Sigma A_i$$
$$= C^* + \Sigma \operatorname{tr}[B_i \operatorname{var} u_i],$$

where

$$\operatorname{tr}[B_i \operatorname{var} u_i] = E\{[u_i - u_i^*]'B_i[u_i - u_i^*]\}.$$

The solution x_i^n corresponds to Martin Weitzman's (1974) quantity solution in his treatise on price vs. quantity. We are interested in two control mechanisms whose cost performances lie somewhere between the perfect control case and the no *ex post* information case.

A. *Bounded Rational Control*

Let us consider a control mechanism in which the managing unit sends assignments to each shop. Management has perfect a priori knowledge concerning technologies of shops (i.e., it knows B_i's) and is capable of calculating the optimal solution once an emergent event is identified, but its rationality is bounded in two respects: it can identify emergent events only imprecisely, and can have production assignments implemented only with some delay—meanwhile, events may change.

First, suppose that the perception u_i^{br} of management of an emergent event u_i is errored in such a way that

$$u_i^{br} = u_i + v_i \qquad (i = 1, \dots, n)$$

where v_i is a stochastic vector with $E[v_i] = 0$ and $E[v_i v_i'] = \gamma E[(u_i - u_i^*)(u_i - u_i^*)']$ for some constant γ. Given the quadratic loss function, expected loss is minimized by setting $u_i = u_i^{br}$ in equation (1). Production assignments would therefore be

$$x_i^{br} = u_i^{br} + B_i^{-1}B[x^* - \Sigma u_i^{br}] \qquad (i = 1, \dots, n).$$

Second, it is assumed that time Δ elapses between the perception of an emergent event by the management and the implementation of the corresponding production assignments and that emergent events constitute a first-order autoregressive process defined by

$$du_i(t) + hu_i(t)\,dt = d\mu_i(t) \qquad (i = 1, \dots, n),$$

where h is a positive parameter representing the rate at which the environment forgets the past history of u_i and $\{\mu_i\}$ is a random process with orthogonal increments. It can be shown that the process $u_i(t)$ defined by the above equation satisfies the following relation:[10]

$$E\{[u_i(t) - u_i^*][u_i(t+\Delta) - u_i^*]'\} = e^{-h\Delta} \operatorname{var} u_i(t).$$

When an event $u_i^{br}(t)$ $(i = 1, \dots, n)$ is perceived by management and the corresponding production assignments are implemented after time Δ, the expected instantaneous total cost at the time of implementation would be

$$C^{br} = \Sigma[x_i^{br}(t) - u_i(t+\Delta)]'B_i \times [x_i^{br}(t) - u_i(t+\Delta)] + \Sigma A_i$$
$$= C^* + \operatorname{tr}[B \operatorname{var} U] + (\gamma + 2(1 - e^{-h\Delta})) \times \{\Sigma \operatorname{tr}[B_i \operatorname{var} u_i] - \operatorname{tr}[B \operatorname{var} U]\}$$
$$= C^n + (\gamma + 1 - 2e^{-h\Delta})(C^n - C^p).$$

Actually the production assignment x_i^{br} $(i = 1, \dots, n)$ does not minimize the expected cost given the perception error and the time lag in response to emergent events. This can be easily seen from the observation that if γ or $h\Delta$ is sufficiently large, C^{br} can exceed the cost C^n under no *ex post* information. In such cases it is better not to react to changing events, but to make routine the noninformation solution (i.e., the Weitzman solution).

---

[10] See A. M. Yaglom (1962, chs. 2 and 3).

If management is rational enough to recognize its bounded rationality, what it ought to do to reduce the expected cost is to mix the production assignments x_i^n and x_i^{br}.[11] In other words, before utilizing perceived information u_i^{br}, management ought to shrink it to the expected value. Let

$$x_i^r(t+\Delta) = \alpha x_i^{br}(t) + (1-\alpha)x_i^n \qquad (i=1,\dots,n).$$

Then the aggregate instantaneous cost would be

$$C^r = C^n + \{(1+\gamma)\alpha^2 - 2\alpha e^{-h\Delta}\}(C^n - C^p).$$

By choosing $\alpha = e^{-h\Delta}/(1+\gamma)$, the expected cost would be minimized at the level

$$C^r = C^n - e^{-2h\Delta}(C^n - C^p)/(1+\gamma).$$

Suppose that the time discount rate of the system is δ. (It is assumed that the rate of discount captures the effect of technological obsolescence as well.) Then the sum of discounted expected costs from the present to the infinite horizon would be

$$\int_0^\infty e^{-\delta t} C^r dt$$

$$= (1/\delta)\{C^n - e^{-2h\Delta}(C^n - C^p)/(1+\gamma)\}.$$

B. *Learning by Doing*

An alternative mechanism in controlling the system is the horizontal coordination in which decisions on inputs and outputs of shops are coordinated among relevant shops. But the shops capacity to coördinate is limited in two respects: the ability to identify emergent events is initially limited and can only be improved upon over time through learning by doing; and the horizontal coordination by shops is imperfect as each shop lacks information concerning technologies of other shops except for those directly related. However, once shops acquire the skill to identify emergent events affecting their own and their transaction partners' cost conditions, they can implement (at least locally) an appropriate production plan without delay. First, let us concentrate upon the effect of learning by doing by assuming that shops can somehow find and implement the optimal solution of production corresponding to a perceived emergent event. Let us call this mechanism—in which the perceptive abilities of shops improve over time and they become perfectly capable of calculating optimal solutions—quasi-horizontal coordination. The problems arising from imperfect coordination due to the lack of centralized technological information will be considered later.

Assume that the shops' perceptive abilities are represented by the probability f of identifying true emergent events. Suppose that, if the shops identify a situation, an optimal solution correspondent to it is implemented. If not, the no *ex post* information solution x_i^n $(i=1,\dots,n)$ is routinized. Then the instantaneous expected total cost of the systme under this quasi-horizontal coordination would be

$$C^{qh} = C^n - f(C^n - C^p).$$

Suppose that the probability of problem recognition improves over time through learning by doing according to

$$df(t)/dt = kf(t)\{\log[1/f(t)]\},$$

where k is a positive parameter representing the rate of learning. Upon integration, we obtain the Gomertz growth model

$$f(t) = \exp\{-\beta e^{-kt}\}.$$

When $t=0$, $f(0)=e^{-\beta}$ so that the inverse of β may be identified with the initial ability of shops as regards to problem identification. As t goes infinity, f approaches one.

The sum of expected discounted costs under the quasi-horizontal coordination from the present to the infinite horizon is

---

[11] Since I am concerned with a quadratic programming, the optimal solution is a linear function of x_i's.

$$\int_0^\infty e^{-\delta t} C^{qh}(t)\,dt = (1/\delta)C^n - \int_0^\infty \exp(-\delta t - \beta e^{-kt})\,dt(C^n - C^p)$$

$$= (1/\delta)C^n - (1/k)\int_0^\infty \exp\{-(\delta/k)y - \beta e^{-y}\}\,dy(C^n - C^p).$$

By setting $\delta/k = \theta$

$$= (1/\delta)\{C^n - \theta\beta^{-\theta}\nu(\theta,\beta)(C^n - C^p)\},$$

where ν is the incomplete *gamma* function:[12]

$$\nu(\theta,\beta) = \int_0^\beta e^{-s} s^{\theta-1}\,ds.$$

The value of $\theta\beta^{-\theta}\nu(\theta,\beta)$ is monotonically decreasing in both θ and β. By comparing the sum of expected discounted costs under the bounded rational control and the quasi-horizontal control, we have the following proposition.

PROPOSITION 1: *The sum of discounted expected costs under quasi-horizontal coordination is smaller than that under bounded rational control if and only if*

$$e^{-2h\Delta}/(1+\gamma) < \theta\beta^{-\theta}\nu(\theta,\beta).$$

Thus the relative advantage of quasi-horizontal coordination is positively related to the lag in hierarchical adaptation to emergent new events represented by Δ, to the degree of imprecision of management's perception represented by γ, to the relative ratio of the rate of learning of shops to the time discount rate represented by θ, and to the initial ability of shops to perceive emergent events represented by the inverse of β. This is independent of technological data represented by B_i's and var u_i's. Specifically, if $\theta = 1$, then $\nu(1,\beta) = 1 - e^{-\beta}$ so that the condition can be simply written as

$$e^{-2h\Delta}/(1+\gamma) < \beta^{-1}[1 - f(0)].$$

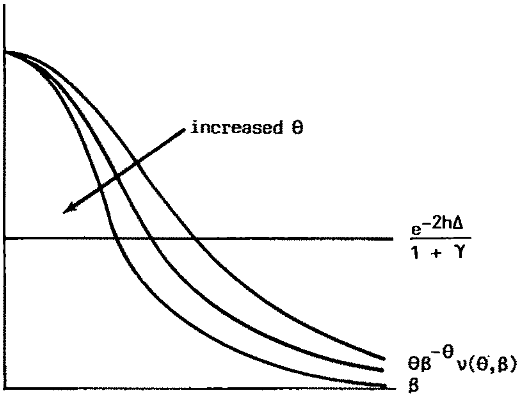


FIGURE 1

Since $\lim_{\beta\to 0}\theta\beta^{-\theta}\nu(\theta,\beta) = 1$ and $\lim_{\beta\to\infty}\theta\beta^{-\theta}\nu(\theta,\beta) = 0$, it holds that

COROLLARY 1: *The value of β must be smaller than a certain strictly positive maximum value in order for the quasi-horizontal coordination to be relatively advantageous. This maximum value of β decreases as h, γ, and Δ decrease and as θ increases. (See Figure 1.)*

This corollary suggests that in order for the quasi-horizontal coordination to be relatively advantageous, the initial ability of shops to identify emergent events must be higher than a certain level. Even if the rate of learning is very high relative to the time discount rate, learning alone cannot make the quasi-horizontal coordination a superior mechanism.

C. *Imperfect Horizontal Coordination*

Let us now take into account the imperfect ability of shops to coordinate their production decisions. Suppose that transactional decisions concerning any products are coordinated between the relevant supplier shop(s) and the relevant user shop(s). Suppose that, with probability f evolving as already specified, they are able to mutually identify emergent events affecting each others' cost condition associated with the transaction of that particular product, as well as to share technological knowledge concerning mutual B_i's. When they do so, they agree upon the particular amount of the transaction of that product which minimizes the joint costs, assuming other inputs and outputs are fixed at the solution under no *ex post* information. If they cannot share information, they routinize the transaction at the level of the no *ex post* information solution. For example, suppose that product k is transacted between shop i and shop j. Further, suppose that it is only these two shops that either produce or use the product k. Then shop i and shop j, mutually recognizing the true value of u_{ik} and u_{jk} with probability f, choose x_{ik} and x_{jk} which would minimize the joint costs subject to the constraint $x_{ik}+x_{jk}=x_k^*$, assuming that other components of x_i and x_j are fixed at the level specified by x_i^n and x_j^n and those of u_i and u_j are fixed at the level specified by u_i^* and u_j^*.

---

[12] See I. S. Gradshteyn and I. M. Ryzhik (1980, pp. 308, 940).

In general, the local joint cost minimizing, when *ex post* information is shared by relevant shops, requires that

$$B_i(x_i^n-u_i^*)+\hat{B}_i[(x_i-u_i)-(x_i^n-u_i^*)]=\lambda \quad (i=1,\ldots,n)$$

and $\Sigma x_i = x^*$ for some intrafirm shadow price λ, where

$$\hat{B}_i=\begin{bmatrix} b_{i,11} & & & & 0 \\ & \ddots & & & \\ & & b_{i,jj} & & \\ & & & \ddots & \\ 0 & & & & b_{i,ss} \end{bmatrix}.$$

From this, the imperfectly coordinated solution x_i^{ip} satisfies

$$\begin{aligned} x_i^{ip} &= u_i+B_i^{-1}B[x^*-U^*]-\hat{B}_i^{-1}\hat{B}[U-U^*] \\ &= x^p+[B_i^{-1}B-\hat{B}_i^{-1}\hat{B}][U-U^*] \\ & \qquad (i=1,\ldots,n), \end{aligned}$$

where $\hat{B}_i^{-1}$ is a diagonal matrix with the jth element being $1/b_{i,jj}$ if $b_{i,jj}$ is positive and zero otherwise. Then the total instantaneous expected cost under the imperfect horizontal coordination when the true event is perceived is

$$\begin{aligned} C^{ip} &= C^p+\mathrm{tr}[(\Sigma\hat{B}\hat{B}_i^{-1}B_i\hat{B}_i^{-1}\hat{B}-B)\mathrm{var}\,U] \\ &= C^n-\Sigma\{\mathrm{tr}[(B_i-\hat{B}\hat{B}_i^{-1}B_i\hat{B}_i^{-1}\hat{B})\mathrm{var}\,u_i]\}. \end{aligned}$$

One can define efficiency ratio of the imperfect horizontal coordination relative to perfect control by

$$\begin{aligned} \eta(B) &= (C^n-C^{ip})/(C^n-C^p) \\ &= \frac{\Sigma\{\mathrm{tr}[(B_i-\hat{B}\hat{B}_i^{-1}B_i\hat{B}_i^{-1}\hat{B})\mathrm{var}\,u_i]\}}{\Sigma\{\mathrm{tr}[(B_i-B)\mathrm{var}\,u_i]\}} \end{aligned}$$

Following the same reasoning as in the case of the quasi-horizontal coordination, we can derive

PROPOSITION 2: *The imperfect horizontal coordination based upon learning is superior to the bounded rational control if and only if*

$$e^{-2h\Delta}/(1+\gamma)<\theta\beta^{-\theta}\nu(\theta,\beta)\eta(B).$$

As all off-diagonal elements of B_i ($i=1,\ldots,n$) approach zero, $\eta(B)$ approaches one. Therefore, somewhat loosely we can claim

COROLLARY 2: *Given other parameters, the relative advantage of the imperfect horizontal coordination would increase vis-à-vis the bounded rational control, as the technological matricies, B_i, become more strongly dominant-diagonal.*

Roughly speaking, the more strongly dominated the marginal cost of each output is by the output level of that product than anything else, the more the relative advantage of the imperfect horizontal coordination would increase vis-à-vis the bounded rational control. This means that if the technology is characterized by near-constant returns to scale, the imperfect horizontal coordination would not perform well. On the other hand, it would perform relatively well when there is a separate capacity constraint for each output.

III. Concluding Remarks

Section II above formalizes the intuitive notion that decision-making power in the Japanese industrial organization is, as far as operational decisions are concerned, relatively widespread. This "decentralized" tendency is to use the on-the-spot knowledge of subunits (workers, in-house shops, and relatively smaller subcontractors). Also, in order that the localized use of on-the-spot knowledge would not lead to inefficient strategic haggling between subunits, sharing of knowledge among neighboring subunits (among workers, among shops, and between prime contractors and subcontractors) is emphasized. The accumulation of on-the-spot knowledge as well as sharing of knowledge can be, needless to say, fostered only over time. That is one of the primary reasons why the Japanese tend to emphasize long-term relations such as "lifetime" employment and stable relational contracts with subcontractors.

On the other hand, emphasis on specialization in the United States seems to have two consequences: First, as jobs become more specialized and standardized, a separate market for each job tends to develop outside the firm, facilitating interfirm mobility of workers. Second, in order to integrate specialized and compartmentalized jobs within the firm, the integration itself has become a specialized function of management. Thus American industrial organization is characterized relatively more by interfirm mobility of workers cum vertical control, whereas Japanese industrial organization is characterized relatively more by intrafirm mobility (rotation) cum horizontal coordination. Needless to say, this difference is only a matter of relative degree and should not be taken as an absolute principle.[13] I simply suggest that there may be a close connection between labor market characteristics and information systematic characteristics of the firm from a comparative perspective.

How does the relative difference between the *A* firm and the *J* firm in their approach toward coordination of operational decisions interact with the mode and direction of more strategic decision making such as research and development? This is an interesting question, however, it is beyond the scope of this paper. I believe that the distinction of the two types of knowledge emphasized in their respective approaches may have some bearing upon this question.

As Hayek (1945) emphasized in the context of the price mechanism vs. central planning, economically useful knowledge may not be limited to that which can be described in terms of formal language (for example, blueprints, books, patents), but may also include intangible skills, undefined tacit knowledge, etc., as generated and transmitted through experiences and on-the-spot contacts. The Japanese firm that relies more on collective learning by doing and informal knowledge sharing in its operations seems to place relative emphasis in research and development efforts on systematizing and developing potentially useful knowledge of the second kind, collectively accumulated within the firm through production experiences. Even when new knowledge is acquired from outside the firm in the form of patents and the like, that new knowledge is often reinterpreted and developed in reference to the firms' own productive experiences. For example, some of the firms most active in the development of biotechnology in Japan are traditional food-processing firms that strive for possible adaptation of their amino acid fermentation and/or brewing techniques to screening and breeding of new micro organisms that have been genetically modified by the use of recombinant DNA. On the other hand, the American firm based on professional specialization and managerial vertical control in its operations seems to place relative emphasis on the scientific efforts of professionally trained researchers in *R&D* under the entreprenurial direction of top management.

---

[13] This contrast of interfirm mobility vs. intrafirm mobility between American industry and Japanese industry is well described in Cole (1979). See also Masanori Hashimoto and John Raisian (1985).

Of course, when the Japanese firm systematizes and develops its own production experiences, the infusion of researchers' and engineers' expertise is indispensable. But even for researchers and engineers, on-the-job training is regarded as equally important as formal training. Therefore, quitting by researchers and engineers in midcareer is discouraged by a system of seniority salaries and retirement benefits contingent on lifetime employment. As a consequence, the rate of interfirm mobility is relatively low even for researchers and engineers. Some argue that the low rate of interfirm mobility of researchers deters rapid technological diffusion without possible government assistance,[14] but one may counterargue that the lower exit rate of researchers reduces the risk of premature leakage of research results out of the firm, thus inducing the firm to finance *R&D* more actively without fearing the free riding of quitters.

Professional specialization in *R&D* in the United States makes professional communications across the boundary of firms, as well as interfirm mobility of researchers, relatively more frequent. Also, one of the possible implications of the vertical control approach is the ease of adding new subunits to a system, or starting an entirely new system, by managerial initiatives. The relatively high mobility of researchers combined with the relative ease of structural reorganization makes it easier to commercialize new knowledge in the form of acquisition of small innovative firms by large established firms or spinoffs of venture businesses. In contrast, the internal diversification or spinoffs of semiautonomous subsidiaries by established firms is the more dominant form of commercializing new knowledge in Japan.

Again I emphasize that the difference in these two approaches to the generation, development, and commercialization of new knowledge is simply a matter of relative degree, but I submit that it is not reasonable to neglect the subtle difference as irrelevant for comparative studies of industrial organization between the two economies. Conventional wisdom may dictate, then, that the American approach is more conducive to scientific research which may lead to a true breakthrough in technology, whereas the Japanese approach is more consistent with applied research and product development. But whether this conventional wisdom is an empirically supportable proposition or not is yet to be seen.

REFERENCES

Aoki, Masahiko, (1984a) *The Economic Analysis of the Japanese Firm*, Amsterdam: North-Holland, 1984.

______, (1984b) "Innovative Adaptation through the Quasi-tree Structure: An Emerging Aspect of the Japanese Entrepreneurship," in D. Bös et al, eds., *Entreprenurship, Zeitscrift fur National Ökonomie*, 1984, supplementum 4, 177–98.

______, "Incentives to Share Knowledge and Risk," presented to Arne Ryde Symposium on Incentive Problems in Some Major Economic Systems, University of Lund, August 1985.

Arrow, Kenneth J., "Vertical Integration and Communication," *Bell Journal of Economics*, Spring 1976, *6*, 173–83.

______, "The Information Structure of the Firm," *American Economic Review Proceedings*, May 1985, *75*, 303–07.

Asanuma, Banri, "The Organization of Parts Purchases in the Japanese Automotive Industry," *Japanese Economic Studies*, Summer 1985, *9*, 32–53.

Coase, Ronald, "The Nature of the Firm," *Economica*, November 1937, *4*, 386–405.

Cole, Robert E., *Work, Mobility, and Participation: A Comparative Study of American and Japanese Industry*, Berkeley: University of California Press, 1979.

______, "Target Information for Competitive Performances," *Harvard Business Review*, May–June 1985, *63*, 106–09.

______ **and Yakushiji, Taizo,** *The American and Japanese Auto Industries in Transition*, Ann Arbor: University of Michigan, 1984.

Cramer, Jacque, "A Partial Theory of the Optimal Organization of a Bureaucracy," *Bell*

---

[14] Such an opinion is expressed and the following counterpoint is discussed by Gary Saxonhouse (1984).

Journal of Economics, Autumn 1980, *11*, 683–93.

Geanakopolos, John and Milgrom, Paul, "A Theory of Hierarchies based on Limited Managerial Attention," unpublished manuscript, 1985.

Gradshteyn, I. S. and Ryzhik, I. M., *Table of Integrals, Series, and Products*, New York: Academic Press, 1980.

Grossman, Sanford and Hart, Oliver, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, forthcoming.

Hashimoto, Masanori and Raisian, John, "Employment Tenure and Earnings Profiles in Japan and the United States," *American Economic Review*, September 1985, *75*, 721–35.

Hayek, Friedrich, "The Use of Knowledge in Society," *American Economic Review*, September 1945, *35*, 519–30.

Klein, Benjamin, Crawford, Robert A. and Alchian, Armen A., "Vertical Integration, Appropriable Quasi-rents and the Competitive Contracting Process," *Journal of Law and Economics*, October 1978, *21*, 297–326.

Koike, Kazuo, "Skill Formation Systems in the U.S. and Japan: A Comparative Study," in M. Aoki, ed., *The Economic Analysis of the Japanese Firm*, Amsterdam: North-Holland, 1984, 47–76.

______, "Human Resource Development in the Japanese Industry," mimeo., 1985.

Kreps, David, "Corporate Culture and Economic Theory," mimeo., Stanford University, 1984.

Monden, Yasuhiro, *Toyota Production System*, Atlanta: Industrial Engineering and Management Press, Institute of Industrial Engineers, 1983.

Saxonhouse, Gary, "Industrial Policy and Factor Markets: Biotechnology in Japan and the United States," mimeo., University of Michigan, 1984.

Weitzman, Martin, "Prices and Quantities," *Review of Economic Studies*, October 1974, *44*, 511–18.

Williamson, Oliver E., *Markets and Hierarchies: Antitrust Implication*, New York: Free Press, 1975.

______, *The Economic Institutions of Capitalism*, New York: Free Press, 1985.

Yaglom, A. M., *An Introduction to the Theory of Stationary Random Functions*, Englewood Cliffs: Prentice-Hall, 1962.

Technological Opportunity and Spillovers of *R&D*: Evidence from Firms' Patents, Profits, and Market Value

By ADAM B. JAFFE*

This paper quantifies the effects on the productivity of firms' R&D of exogenous variations in the state of technology (technological opportunity) and of the R&D of other firms (spillovers of R&D). The R&D productivity is increased by the R&D of "technological neighbors," though neighbors' R&D lowers the profits and market value of low-R&D-intensity firms. Firms are shown to adjust the technological composition of their R&D in response to technological opportunity.

The determinants of "the rate and direction of inventive activity" (Richard Nelson, 1962) have been the focus of extensive theoretical and empirical work. This paper presents quantitative estimates of the effects on the productivity of private research and development (*R&D*) of two factors from the "supply side" of innovation: "technological opportunity" and *R&D* "spillovers." I define technology opportunity as exogenous variations in the cost and difficulty of innovation in different technological areas. These variations may be due to intrinsic characteristics of the technology, or they may be due to the state of exogenous scientific knowledge at a particular point in time. Therefore, the pattern of technological opportunity may change over time, though I assume that it generally changes slowly, requiring a number of years for significant changes to manifest themselves. F. M. Scherer (1965,1982) has investigated the importance of technological opportunity by estimating the importance of industry dummies in explaining innovation rates. This paper refines this approach by using more detailed data to characterize firms technologically, and by employing multiple indicators of innovative success.

The other influence on the firms' *R&D* program that I examine is spillovers of *R&D* from other firms (Zvi Griliches, 1979). Since knowledge is inherently a public good, the existence of technologically related research efforts of other firms may allow a given firm to achieve results with less research effort than otherwise. Edwin Mansfield (1977) has found that, for certain chemical innovations, imitation is much cheaper than innovation, and private rates of return are only a fraction of social rates of return. This paper attempts to gauge the broad importance of the spillover phenomenon by looking at the average effect that other firms' *R&D* has on the productivity of a firm's own *R&D*. This indirect form of inference is made necessary by the extreme difficulty of observing the actual spillovers except anecdotally. My case for the importance of spillovers is, therefore, entirely circumstantial, but I believe that the consistent evidence from three distinct indicators of innovative success makes the case credible.

From a purely *technological* point of view, *R&D* spillovers constitute an unambiguous positive externality. Unfortunately, we can only observe various *economic* manifestations of the firm's *R&D* success. For this reason, the positive technological externality is potentially confounded with a negative effect of others' research due to competition. It is not possible, with available data, to distinguish these two effects; I do, however, find evidence that both are present.

It should be emphasized that I do not model the strategic interaction of the firms doing *R&D* in similar areas, with or without spillovers. I assume throughout that firms hold Cournot conjectures about the effects of their actions on the other firms, and I do not in any way impose equilibrium on the actions of the different firms.

---

*Harvard University, Cambridge, MA 02139. This paper is based on my Ph.D. dissertation (1985). I am indebted to Zvi Griliches, Richard Caves, Mark Schankerman, two referees, and workshop participants at Harvard and the NBER for many helpful comments. Financial support was provided by the National Science Foundation under grant PRA81-08635 and by an Alfred P. Sloan Doctoral Dissertation Fellowship.

The spillover and technological opportunity phenomena have in common that their effect on a particular firm will depend on the technological nature of the firm's research. My dissertation develops a methodology for characterizing the technological position of the firm's research program and constructing variables related to spillovers and technological opportunity. This methodology is summarized in Section I; Section II develops the estimating equations. Section III presents the main empirical results and Section IV presents additional results examining the changes in the apparent pattern of technological opportunity between the two periods.

I. Characterizing the Technological Position of Firms

In order to look for the effects of technological opportunity and spillovers, we should identify the technological areas in which the firms are engaged in research. One could imagine, if there are K such areas that relate to manufacturing products or processes, that the "technological position" of a firm's research program could be characterized by a vector $F=(F_1 \dots F_K)$ where F_k is the fraction of the firm's research budget devoted to area k.[1] This, unfortunately, is not observable, but data do exist that are closely related. The patents received by the firm are classified by the patent office into technology-based patent classes.

This classification system provides the basis for the search of "prior art" necessary before a new patent can issue. It is important to note that this system is *not* an alternative product or industry classification. It is technology based. There is surely some relationship between industries and patent classes, in the sense that firms in a given industry will, on average, patent more in some classes than in others. But the mapping from classes to industries is not unique in either direction. This is illustrated by an example given by Jacob Schmookler, who used these data extensively. He notes that in a patent subclass relating to "dispensing of solids" (1966, p. 20), he found patents for a toothpaste tube and a manure spreader.

The fact that the patent classification system is technology based rather than product based is often viewed as a limitation on the economic usefulness of these data. It probably is so if one is seeking a measure of innovation by industry, for use in, for example, a structure-performance study. For my purposes, however, it is a decided virtue.

I utilize the distribution of the firms' patents over patent classes to characterize the technological position of the firm. In the basic data, there are 328 patent classes (though the classification system actually contains thousands of subclasses). This information is available for the patents granted between 1969 and 1979 to about 1700 manufacturing firms in the *R&D* panel that has been assembled recently at the National Bureau of Economic Research (NBER). This data set is a marriage of Compustat and Patent Office data that are documented in Clint Cummins et al. (1984) and John Bound et al. (1984). The companies in the data set were granted about 260,000 patents over the period. The average firm has patents in about 20 different classes. The classes themselves vary greatly in importance, from "Chemistry, carbon compounds" with 20,000 patents taken by 340 different firms to "Bee culture" which has one patent.[2] To make the distribution vectors empirically usable, the 328 classes were grouped into 49 categories. This grouping was essentially *ad hoc*, based on the names, with more aggregation of classes with few patents and less aggregation of those with many.

---

[1] One might ask whether it is the technological nature of the firm's research program or that of its current production technology that really matters. To the extent that the two do differ, it would seem that what matters most for the effects of technological opportunity and spillovers is the nature of the research program.

[2] There are 8 classes in which these firms took no patents, including "Land vehicles, animal draft" and "Whips and whip apparatus."

Firms' technological positions are, in the long run, a matter of choice for the firms. If technological opportunity affects profitability, then we would expect firms to move to the more profitable positions. I expect, however, that changes in the technological position of firms can be brought about only slowly. Expertise in various areas is not easily acquired, and goodwill and reputation in product markets represent sunk costs that make jumping costly. In this study, reflecting the view that technological position is endogenous, but only over relatively long time periods, I construct two technological position vectors for each firm: one is based on all patents applied for up to 1972, and one is based on all patents applied for after 1972.[3] In the estimation, I treat each of these distributions as exogenous for the purpose of regressions involving cross sections of firms at the end of each period. I then examine the changes in position between the two periods, to determine whether they can be explained as responses to the pattern of technological opportunity.

Thus the basic data characterizing the firm's technological position are two 49-element distribution vectors, one for the earlier period and one for the later. These vectors are used in two ways. First, they are utilized to construct a variable that will be used to infer the existence of spillovers. I assume that the existence of technological spillovers implies that a firm's *R&D* success is affected by the research activity of its neighbors in technology space. To make this notion operational requires significant additional structure. I assume that the total relevant activity of other firms can be summarized by a "potential spillover pool" that is simply a weighted sum of other firms' *R&D*, with weights proportional to the proximity of the firms in technology space.[4] To measure the proximity of firms i and j, I use the angular separation or uncentered correlation of the vectors F_i and F_j:

$$P_{ij} = F_i F_j' / \left[(F_i F_j)(F_{ji}) \right]^{1/2}. \tag{1}$$

This proximity measure has the following properties: it is unity for firms whose position vectors are identical, it is zero for firms whose vectors are orthogonal, and it is bounded between 0 and 1 for all other pairs. It is closer to unity the greater the degree of overlap of the two firms' research interests.[5]

The potential spillover pool S_i is constructed using the proximities defined in equation (1) as weights in a summation of all other firms' *R&D* spending, denoted R_j:

$$S_i = \sum_{j \neq i} P_{ij} R_j. \tag{2}$$

---

[3] Classifying the patents by date of application rather than date of issue is preferable because that is when the firm perceived itself to have generated new knowledge, and because there are long and variable lags in the Patent Office's Processing of Applications. Recall, however, that the data set consists of all patents *granted* to these firms between 1969 and 1979. Because of the lags in granting, this includes essentially all of the successful applications filed between 1968 and 1976, but only some of those filed 1977–79, and also includes some from 1965–67 and a sprinkling from before. If we can view the patents we do have for the incomplete years as a random sample of all applications, then the incompleteness doesn't matter for this purpose. If, however, some patent classes have systematically long or short processing lags, then the early period totals are biased towards long-lag categories and the later period towards short-lag categories.

[4] This is the basic approach suggested by Griliches (1979). Jeffrey Bernstein and M. Ishaq Nadiri (1983) measure spillovers using a cost function approach and time-series data for the chemical industry. Their pool variable is the unweighted sum of the *R&D* of all other firms in the industry.

[5] This measure of proximity is purely directional; i.e., it is not directly affected by the length of the F vectors. The length of the vector depends on the degree of focus or concentration of the firm's research interests. (The length is actually the square root of the Herfindahl index of concentration of the category shares.) Other proximity measures, notably the Euclidean distance between the vector endpoints, are very sensitive to the length. For example, all pairs of relatively diversified firms are "close" by the Euclidean measure, because they are close to the origin of the coordinate system, even if the vectors are orthogonal. If the vectors are all normalized to have the same length, then $(1-P_{ij})$ is proportional to the squared Euclidean distance between the endpoints of the vectors F_i and F_j.

TABLE 1—STATISTICS FOR TECHNOLOGICAL CLUSTERS

			1972 Data					
	Number of Firms		Sales		*R&D*		Patents	
Cluster	1972	1978	Mean	Median	Mean	Median	Mean	Median
Adhesives & Coatings	30	24	369	190	3.2	1.9	6.7	2
Chemistry, Carbon	45	39	1064	619	28.2	11.3	66.8	24
Chemistry, Electrochemistry	16	13	708	189	44.8	12.6	89.3	33
Chemistry, Organic	21	22	565	402	26.3	19.5	45.7	28
Cleaning and Abrading	16	14	506	312	7.9	3.0	19.2	13
Compositions	20	20	3326	1890	25.6	17.1	86.6	73
Cutting	24	15	131	90	2.2	1.6	6.2	3
Elec Computers & Data Proc.	21	20	997	188	61.5	8.1	63.7	13
Elec Transmission & Systems	34	26	640	104	18.9	1.5	57.2	6
Electronic Communication	28	32	923	251	24.1	8.5	53.9	17
Fluid Handling	27	24	613	213	12.2	3.4	40.0	7
Food	34	29	883	684	5.4	3.0	7.6	3
Measuring, Testing & Signaling	31	20	380	150	7.3	2.6	14.2	9
Medical	13	25	323	108	8.9	2.5	17.1	10
Metals and Metal Working	33	29	964	530	12.4	4.1	21.5	12
Misc Consumer Goods	24	29	335	165	4.6	4.2	11.9	5
Power Plants (Non-Electric)	51	29	1758	343	56.7	6.9	48.5	13
Receptacles & Packages	24	36	352	150	3.6	1.0	11.8	4
Refridgeration & Heat Exch.	29	37	367	220	5.2	2.3	13.3	6
Static Structures	27	38	199	104	1.1	.7	2.7	2
Vehicles	25	36	440	168	10.3	2.6	15.8	7
All Firms	573	557	801	236	19.0	3.4	34.0	8

Note: Sales and *R&D* are in millions of dollars.

This approach assumes that the conditions of appropriability are the same across all technological areas. Suppose, as in the formulation of A. Michael Spence (1984), that imperfect appropriability means that a fraction θ of each firm's research "leaks out." If $\theta = 0$, then appropriability is perfect; if $\theta = 1$, *R&D* is a pure public good. For my formulation, it doesn't matter what θ is,[6] but I have assumed that it is the same for all firms. To the extent that this is untrue, S_i is poorly measured.

The 49-element technological position vectors were also utilized with a clustering algorithm to assign the firms to "technological groups."[7] The idea is to identify those firms whose technological focus is sufficiently similar that they face the same state of technological opportunity. Initial clusters were formed by taking about 25 of the most populous patent categories and assigning each firm to a cluster corresponding to the category in which that firm has the most patents. Firms were then iteratively reassigned to clusters to maximize the probability of observing the firms' category distributions on the assumption that firms in a given cluster are all drawing patents from the *same* underlying multinomial distributions. As in all clustering procedures, the number of clusters must be determined arbitrarily. It was found that the fit deteriorated rapidly at about 20 clusters, so 21 were used. This clustering was carried out twice, yielding for each firm a technological cluster assignment for 1972 and one for 1978. These assignments can be used to construct dummy variables for the technological groups. These dummy variables will be used in the estimation below to allow for variations in technological opportunity.

---

[6] The units of s_i are arbitrary, because there is no natural normalization of the proximity weights. This does not matter because I am primarily interested in elasticities. Of course, if $\theta = 0$, then this pool is a mere illusion and should have no direct effect on firm i.

[7] The algorithm used was derived from the K-means algorithm (J. B. McQueen, 1967, and J. A. Hartigan, 1975), modified to exploit the multinomial structure of the elements of the F-vectors. For more detail, see my dissertion.

Table 1 summarizes the technological clusters. The name of each cluster derives from the name of the patent class that formed the initial seed for the cluster. Almost one-third of the firms changed clusters between the two periods. Despite this, the clusters themselves were fairly stable; the proximity (as defined by equation (1)) between the *center* of each cluster in the early period, and its center in the later period averaged .96 and was not below .90 for any cluster. In Section IV, I will show that the moves that did occur can be related to the pattern of profitability across the groups.

II. Modeling the Productivity of *R&D*

The general approach taken here is to assume that there exists a stable relationship between the investment by the firm in *R&D* and the production of new economically useful knowledge. This relationship is conditioned by the spillovers that the firm receives and the state of technological opportunity. We do not observe new knowledge, but we do observe several of its consequences. On average, if *R&D* is a sensible activity for the firm, the new knowledge should lead eventually to the generation of profits. This profit stream should be reflected in the firm's market value. Along the way, the new knowledge may also generate patent applications. In this section, I develop estimating equations that relate the firm's patent applications, profits, and market value to its *R&D*, other attributes, the spillover and technological opportunity variables, and industry variables.

This process is dynamic, and much work has been done on the nature of the lags involved.[8] The data are not rich enough to support a dynamic specification with spillover effects. The equations to be estimated should, therefore, be thought of as long-run or steady-state relationships.

A. *The Patent Equation*

I begin by assuming that the new knowledge produced by the firm in any period is related to its *R&D* in that period according to a modified Cobb-Douglas technology of the form:[9]

$$(3)\quad k_i = \beta_1 r_i + \beta_2 r_i s_i + \gamma_1 s_i + \sum_{c=1}^{21} \delta_{1c} D_{ic} + \varepsilon_{1i}$$

where k_i is the new knowledge generated by firm i, r_i is its *R&D* spending, and s_i is the potential "spillover pool" whose construction is described above. All variables are expressed in logs.[10] The D_{ic}'s are a set of dummy variables for the technological clusters discussed earlier; if technological opportunity is important, the δ_{1c}'s will differ across c. The term ε_{1i} is a random disturbance, assumed to be distributed independently but not necessarily identically across i.[11] It includes unobserved firm-specific attributes that may be correlated with r_i; this raises crucial estimation issues that will be discussed below. Equation (3) implies that the *R&D* of other firms may increase the knowledge output of the firm directly, and may also increase the elasticity with respect to own *R&D*. This means that each firm's elasticity of knowledge production with respect to *R&D* is $\beta_1 + \beta_2 s_i$. To facilitate interpretation, I subtract the mean of the spillover variable before estimation, so the elasticity for a firm with the average spillover pool is just the coefficient β_1. If β_2

---

[8] See Ariel Pakes and Griliches (1984), Pakes (1985), and Bronwyn Hall et al. (1983).

[9] Actually, we expect knowledge production to depend on a distributed lag of *R&D*, but this lag structure is difficult to identify, and most of the weight appears to fall on the contemporaneous *R&D*. See Pakes-Griliches, and Hall et al.

[10] Throughout this paper, lowercase letters represent the logs of variables; β's indicate coefficients of endogenous variables and γ's those of exogenous variables.

[11] If there is covariance across firms in the productivity of *R&D*, then that is precisely what constitutes patterns of technological opportunity. Hence I assume that this will be picked up by the terms $\delta_{1c}D_{ic}$.

and γ_1 are both zero, then equation (3) is a standard "knowledge production function" (Griliches, 1979).

Knowledge is not observed, but I assume that a fraction of it is patented according to

$$(4)\quad P_i = \exp\left[\sum_{c=1}^{21} \alpha_c D_{ic}\right]\left[\exp(\eta_i)\right] K_i.$$

Thus, the ratio P_i/K_i (the propensity to patent) for any firm i depends on its technological position and a random firm-specific component. I will again assume η_i is independent across firms. Note that I rule out by assumption a dependence of the propensity to patent on the potential spillover pool; allowing such a dependence would preclude identification of the spillover effect in the patent equation. Because of the necessity of this assumption, the patent equation alone would provide questionable evidence of a real spillover effect. In conjunction with the other results, however, it is a useful part of the story.

Taking logs in equation (4) and substituting into equation (3) yields an equation relating patents to *R&D* and the pool:[12]

$$(5)\quad p_i = \beta_1 r_i + \beta_2 r_i s_i + \gamma_1 s_i + \sum_{c=1}^{21} (\delta_{1c} - \alpha_c) D_{ic} + \varepsilon_{1'i}.$$

Because I allow for the dependence of the propensity to patent on technological position, the technological cluster dummies in the patent equation cannot be related solely to technological opportunity. As will be discussed below, this hinders somewhat the ability to draw conclusions about technological opportunity.

B. *The Profit Equation*

In modeling the profits of the firm, I want to allow for the existence of unobserved firm-specific entrepreneurial skill, as well as possible effects of market power. The modeling approach is to treat market power and the unobserved firm attributes (as well as the *R&D* of other firms) as intangible assets that augment the productivity of the firm's *R&D* and capital stocks. One way to do this is to postulate a modified restricted profit function in log-log form:[13]

$$(6)\quad \pi = \beta_3 m_i + \beta_4 \underline{r}_i + \beta_5 s_i \underline{r}_i + \beta_6 c_i + \gamma_5 s_i + \gamma_6 cn_i^I + \sum_{c=1}^{21} \delta_{2c} D_{ic} + \varepsilon_{2i}$$

where π_i is the gross operating income of firm i; m_i is its market share; and $\underline{r}_i$ is the stock of accumulated *R&D* investments, distinct from the annual flow used in the patent equation. The variable c_i is the capital stock, cn^I is the four-firm concentration ratio in the firm's industry, and the other variables are as before. Note that I could have started by writing equation (6) in terms of the unobserved knowledge stock, and then substituted from the knowledge production function as I did for the patent equation. For patents, this was instructive because of the issue of the propensity to patent. In this case, however, no insight is gained so the intermediate step is excluded.

If $\beta_3 = \gamma_6 = 0$, then equation (6) would be the restricted profit function derived from the production function:

$$(7)\quad q_i = b_0 1_i + b_1 c_i + b_2 \underline{r}_i + b_3 \underline{r}_i s_i + b_4 s_i + \sum_{21}^{c=1} \delta_{2c} D_{ic} + \varepsilon_{2i}$$

by solving out labor through its optimal choice. This would be valid under the assumption that input and output markets are competitive, and that all firms face the same prices or else prices are independent of the exogenous variables and can be absorbed in ε_2. Adding market share and concentration as in (6) has the following interpretation: If output markets are not competitive, then the effect on the firm's revenues can be separated into a component that depends on the firm's own market share, and a component that depends on the overall structure of competition in the industry. Further, these effects enter in a simple multiplicative way.[14]

---

[12] There has been a long debate on the general question of the usefulness of patent data as an indicator of inventive output. For present purposes, suffice it to say that patents have repeatedly passed tests of their economic relevance. See Schmookler, Pakes and Griliches, Bound et al., Pakes, and Mark Hirschey (1982).

[13] For a related approach, see Alan Siu (1984).

C. *The Market Value Equation*

Much attention has been focused in recent years on the conceptual problems of using accounting profits to measure economic profits (Franklin Fisher and John McGowan 1983). Michael Salinger (1984) argues that the stock market provides a better indicator, because it has fewer accounting biases and captures long-run effects. If, however, we wish to infer rates of return on assets, we still must use accounting-based measures of those assets, so the use of market value data does not solve the Fisher-McGowan problem. I view accounting profits and market value as multiple noisy indicators of economic returns.[15]

My approach to relating market value to the firm's measured assets is similar to that of Griliches (1981) and Salinger. The firm's market value is assumed to be a linear function of a weighted sum of the firm's conventional and *R&D* capital stocks:

$$V_i = \bar{q}_i\left[C_i + (\beta_7 + \beta_8 s_i)\underline{R}_i\right]. \tag{8}$$

Equation (8) says that the relative value of a dollar of *R&D* and a dollar of capital depends on the spillover pool; for the firm with the mean pool, the relative value is β_7, which may or may not be unity. The package of capital and *R&D* is valued at a price $\bar{q}$. This price is itself assumed to depend on the firm's market power, its unobserved attributes, and its technological cluster:

$$\bar{q}_i = M_i^{\beta_9} S_i^{\gamma_7} \left(C4_i^I\right)^{\gamma_8} \times \exp\left(\sum_{c=1}^{21} \delta_{3c} D_{ic}\right) \exp(\varepsilon_{3i}). \tag{9}$$

Thus market power, spillovers, and technological opportunity are assumed to augment value multiplicatively rather than additively.[16] This implies that a given degree of market power can be spread over any amount of capital. Compare two firms that have identical shares of identically concentrated industries, but one is twice as big as the other, because its industry is twice as big. It seems plausible that the market value of the larger firm should be twice that of the smaller, as implied by (8) and (9). Salinger assumes that the value is the *sum* of the value of the capital and the value of the market power; this means that the larger firm's value would be less than twice that of the smaller firm.

If we now substitute (9) into (8), divide through by C_i, take logs, and use the approximation $\log(1+x) \approx x$, we obtain

$$v_i - c_i = \beta_7 (R/C)_i + \beta_8 s_i (R/C)_i + \beta_9 m_i + \gamma_7 s_i + \gamma_8 c4_I + \sum_{c=1}^{21} \delta_{3c} D_{ic} + \varepsilon_{3i}. \tag{10}$$

The dependent variable in equation (10) is simply the log of Tobin's q.[17]

---

[14] The concentration and market-share effects are not the focus of this paper, but they are included because they were found to be significant in my dissertation.

[15] For a similar view, see Hirschey and Dean Wichern (1984). Other papers using stock market data to measure profitability include Griliches (1981), Eric Lindenburg and Stephen Ross (1981), and Pakes.

[16] This differs from the treatment in Salinger.

[17] In principle, one could derive the market value directly from the profit equation. This would, however, require solving for the optimal investment streams for capital, *R&D*, and also market power. It is not clear that this is feasible.

TABLE 2—STATISTICS FOR REGRESSION VARIABLES

	Levels (864 Observations)[a]			Log of Levels (864 Observations)		Difference of Logs (1979/1973) (432 Observations)	
	Mean	Median	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Patents	35.5	9.3	120.4	2.30	3.29	−.304	.743
Gross Profits	207.9	32.7	868	4.07	4.62	.243	.475
Tobin's q	1.02	.84	1.80	−.0323	.882	−.142	.512
Annual *R&D*	25.7	3.41	108	1.80	2.98	.177	.562
R&D Stock	121	12.5	98.2	3.51	3.09	.239	.611
Capital Stock	968	153	3562	5.64	6.05	.290	.332
Annual Spillover Pool	2622	2438	2980	7.78	7.80	.152	.344
Spillover Pool Stock	10042	9986	2004	9.45	7.75	.208	.329
Market Share	5.44	3.24	6.23	.848	1.12	–	–
Four-Firm Concentration	38.3	37.1	43.2	3.57	3.62	−.049	.152

Note: All nominal variables are millions of 1972 dollars. Market share and concentration are percentages.

[a] Except for market share, which has only 432.

III. Regression Results

A. *Data and Estimation Issues*

The model is estimated on two cross sections of 432 firms from the NBER *R&D* panel. The firms were chosen on the basis of availability of complete data; the selection is described in more detail in the Appendix. These 432 firms are about one-third of those that reported *R&D* publicly in 1976, but they account for 95 percent of the total reported *R&D* and they do about 80 percent of *all* privately financed *R&D* (see *Research and Development in Industry, 1977,* NSF, 1979). The two cross sections are centered on 1973 and 1979. These years were chosen because they correspond to the approximate end of the periods whose patent data are the basis of the technological position classifications. Each cross section is an average of three years' data, to smooth out transient effects (particularly in profits and patents) and approximate long-run values. Stock variables (capital, *R&D* stock, market value) are measured as of the end of the middle year for each three-year period. All nominal variables are converted to 1972 dollars.

The Appendix describes the data sources and the details of variable construction. Four points should be highlighted. First, I have market share data only for 1972, so I am forced to use this both for 1973 and as a proxy for the 1979 share. Second, the capital stock is a measure of the inflation-adjusted net capital stock of the firm (including an allowance for inventories and unconsolidated subsidiaries). This deals with the most superficial problems of using "book value," but not the deeper criticism of Fisher and McGowan. Third, the patent equation depends on contemporaneous *R&D* spending, but profits and value are assumed to depend on the accumulated *stock* of *R&D*. The *R&D* stocks were constructed from the firm's *R&D* history, assuming 15 percent annual depreciation, and initial stocks fabricated by extrapolating the average growth rate into the infinite past.[18] Finally, all industry-specific variables (market share, concentration, and instrumental variables to be discussed below) are weighted averages for each firm. The weights are the 1972 4-digit SIC sales distributions for each firm. Simple statistics for the regression variables are presented in Table 2.

---

[18] Because most firms do not vary their deflated *R&D* spending very much over time, results are relatively insensitive to different stock constructions.

A long-recognized problem in this sort of estimation is the likelihood that firms differ in ways that we do not observe, such as their management skill. These unobservables are likely to enter positively in the error terms of equations that measure, in any sense, the firm's productivity. They will also influence the investment decisions of optimizing firms. In most simple models, more productive firms would choose higher levels of investment, introducing positive correlation between the error terms and some of the right-hand side variables in the productivity equations, thus biasing the coefficients upward. The classic solution to this problem (if a panel of firms is observed over time) is to assume that the unobservables are unchanging over time for a given firm. This limits their contamination to the cross-sectional dimension of the data, implying that unbiased estimates of structural parameters can be derived from the covariance "within" firms over time. This leads to the so-called "fixed effects" estimates, which are presented below.

This approach has several drawbacks. First, the bulk of the variance in firm data is usually in the cross-sectional dimension. Thus, the fixed effects approach discards much of the information in the data, and often produces uninformative estimates as a result. Second, *R&D* may be measured with error, which leads to bias that is typically *worsened* by within estimation. Finally, the assumption that the unobservables are constant over time is pure convenience. If this is untrue, then the within estimates are also biased.

An alternative is to view the problem as a standard one of endogenous right-hand side variables, and to develop equations for the endogenous variables as a function of other variables. This allows one to use the cross-sectional information, *reduces* bias due to measurement error, and drops the assumption that the unobservables are time-invariant. The price that is paid for these advantages is that one must assume the validity of the instrumental variables. I will devote most of my attention below to estimates in which the firm's *R&D*, capital stock, and market share are treated as endogenous; it is assumed that in the reduced form of the system they depend on industry variables such as the industry's size, growth rate, and *R&D* intensity. The key assumptions implicit in this approach are that the unobserved firm characteristics do not significantly affect these industry variables, and that the industry variables are correctly excluded from the structural equations.

TABLE 3—POOLED *OLS* ESTIMATES FOR 1973 AND 1979 (864 Observations)

	Equation for log of:		
Coefficient	Patents	Profits	Tobin's q
log(*R&D*)	.735	.201	
	(.152)	(.051)	
R&D/Capital			2.95
			(1.52)
log(*R&D*) × log(Pool)	.168	.027	
	(.031)	(.020)	
(*R&D*/Capital) × log(Pool)			.526
			(.192)
log(Pool)	.628	−.078	−.077
	(.108)	(.043)	(.052)
log(Capital)		.564	
		(.024)	
log(72 Share) in 1973		.385	.453
		(.021)	(.043)
log(72 Share) in 1979		.272	.315
		(.021)	(.033)
log(Four-Firm Concentration Ratio)		−.159	−.438
		(.040)	(.052)
F-Statistics for Technological Cluster Dummies (20 *D.F.*):			
1973	4.2	5.8	5.6
1979	3.9	5.7	5.5
R^2	.94	.44	.77
$\hat{\sigma}$	.806	3.46	.509

Note: Each equation includes 21 technological cluster dummies for each year.

B. *The OLS and First-Difference Results*

To provide a reference point for more complicated estimates, I begin by ignoring potential biases and estimating the equations by *OLS*. These results are presented in Table 3. Since this is just a starting point, I comment on the results only generally. First, the

TABLE 4—FIRST DIFFERENCE ESTIMATES FOR 1973 AND 1979
(432 Observations)

	Equation for log of:		
	Patents	Profits	Tobin's q
log($R\&D$)	.398	.098	
	(.225)	(.325)	
$R\&D$/Capital			9.98
			(23.41)
log($R\&D$) ×log(Pool)	.020	.0202	
	(.025)	(.029)	
($R\&D$/Capital) ×log(Pool)			3.34
			(3.03)
log(Pool)	.179	.0179	−.028
	(.058)	(.0583)	(.097)
log(Capital)		−.221	
		(.060)	
log(Four-Firm Concentration Ratio)		−.023	−.259
		(.116)	(.162)
R^2	.038	.101	.031
$\hat{\sigma}$	.728	.450	.504

interaction between own $R\&D$ and the spillover pool is positive; the direct effect of the spillovers is positive in the patent equation but negative in the other two. Second, the technological cluster dummies are extremely significant in all equations. Finally, recall that the basic left-out-variables bias argument is that the coefficients of $R\&D$, capital, and market share in Table 3 are biased upward. This seems plausible for market share, which has large positive and significant effects. It is puzzling, however, that the $R\&D$ coefficient in the patent equation, and the sum of the $R\&D$ and capital coefficients in the profit equation, are significantly less than one, implying strong diminishing returns if these are biased upwards.

Table 4 presents estimates based on the differences between the cross sections; with only two observations per firm these are computationaly equivalent to "within" estimates. Everything is estimated with only limited precision. Implausibly diminishing returns appear in the patent equation. The market value equation coefficients are very large but essentially unidentified. It appears that much of the within variance is noise. We can, perhaps, take an *epsilon* of reassurance from the fact that the results in Table 4 are at least not inconsistent with those in Table 3.

C. *Instrumental Variable Estimates*

The essence of the instrumental variables approach to the problem of unobserved attributes is to model explicitly the dependence of the firm's choice of variables on the exogenous characteristics of its environment. My dissertation presents estimates of a system in which the market share depends on the firm's $R\&D$, industry $R\&D$ and industry sales, and its (optimal) $R\&D$ stock depends on its capital stock, the spillover pool, and industry $R\&D$, sales, and growth rate. This implies that these industry variables are potential instruments for the firm's $R\&D$ and market share. In addition, the industry minimum efficient scale (*MES*) is added as an instrument for the capital stock. The interaction between the firm's own $R\&D$ and the spillover pool makes the equations nonlinear in the variables. In this circumstance there is no optimal nonlinear function of the exogenous variables which should be used as instruments (Takeshi Amemiya, 1983). I utilize the industry variables, the spillover pool, and interactions between the pool and all the industry variables.

The patent, profit, and market value equations for the two different years comprise a six-equation system which can be estimated in a number of different ways. I assume that the 6-element residual vector ε_i is independent of all the instruments, that $E(\varepsilon_i \varepsilon_j') = \underline{0}$ for all $i \neq j$, and that $\mathrm{Var}(\varepsilon_i)$ is an arbitrary matrix Σ_i. Thus I allow for unconstrained covariance across equations and time and conditional heteroskedasticity. This suggests the use of three-stage least squares (*3SLS*), with the procedure described by Halbert White (1982) to calculate standard errors that are robust under conditional heteroskedasticity. Table 5 presents *3SLS* estimates analogous to the *OLS* estimates in Table 3. That is, the slope coefficients (other than for market share) are constrained to be constant over time. Sensitivity of the results to other estimation approaches will be discussed below.

TABLE 5—*3SLS* ESTIMATES OF PATENT, PROFITS, AND *q* EQUATIONS FOR 1973 AND 1979 (432 Observations)

	Equation for log of:		
	Patents	Profits	Tobin's *q*
log(*R&D*)	.875 (.183)	.180 (.042)	
R&D/Capital			3.31 (.209)
log(*R&D*) ×log(Pool)	.352 (.048)	.058 (.020)	
(*R&D*/Capital) ×log(Pool)			.803 (.098)
log(Pool)	.509 (.104)	−.095 (.053)	−.058 (.031)
log(Capital)		.825 (.044)	
log(72 Share) (73 Equation)		.188 (.055)	.310 (.053)
log(72 Share) (79 Equation)		.057 (.055)	.123 (.054)
log(Four-Firm Concentration Ratio)		−.220 (.045)	−.525 (.083)
χ^2_{20} on Technological Cluster Effects			
(1973)	81.3	88.4	98.6
(1979)	94.4	76.5	90.2
$\hat{\sigma}$			
(1973)	.842	3.62	.652
(1979)	.912	3.51	.420

Note: All equations include 21 technological cluster dummies for each year. The instrument set consists of the cluster dummies, the spillover pool flow and the spillover pool stock for both years; 1974 industry *R&D*, 1972 and 1977 industry shipments and concentration, 1972 minimum efficient scale, industry growth rates 1967–72 and 1972–77; and interactions of the pool variables with the contemporaneous industry variables. All instruments are used in all equations. The numbers in parentheses are heteroskedasticity-consistent standard errors calculated according to White (1982).

The results presented in Table 5 are the key results of the paper. Looking first at the patent equation, the estimates imply close to constant returns with respect to the firms' own *R&D*. A firm's *R&D* elasticity is .875 (its coefficient) plus .352 (the *R&D*-pool interaction coefficient) times the firm's log(pool), which is constructed to have zero mean. Thus the elasticity is .875 for the average firm and is higher (lower) for firms with above (below) average pools. The pool has a significant positive effect, both directly and through its influence on the *R&D* elasticity. A firm's pool elasticity is .509 plus .352 times its log(*R&D*). For the firm with mean log(*R&D*), the elasticity of patents with respect to others' *R&D* is about 1.1. If everyone increased their *R&D* by 10 percent, total patents would increase by 20 percent, with more than one-half the increase coming from the spillover effect.

At first glance, this may seem like an implausibly large spillover effect. It is, however, consistent with a reasonable picture of *R&D* as a partial public good. In terms of marginal products evaluated at the mean of the data, the coefficients imply a return of 2.0 patents per million dollars of own *R&D*, and .06 patents per million dollars of other firms' relevant *R&D*. Because the total relevant *R&D* is large relative to any firm's own spending, spillovers are important even though the implied value for Spence's θ is only .03.[19]

The results from the profit and *q* equations are best considered together. Note first that the profit equation estimates imply constant returns in *R&D* and capital for firms with the average spillover pool. The *R&D* share is about 20 percent of the capital share. The elasticity coefficient (for *R&D*, for example) is equal to $(\partial\pi/\partial R)(R/\pi)$. Thus we can estimate the gross rate of return to *R&D* $(\partial\pi/\partial R)$ by multiplying the estimated elasticity by R/π. This yields a gross rate of return for *R&D* of 27 percent; a similar calculation for capital yields 15 percent. These estimates are consistent with the conventional wisdom that the economic depreciation rate for knowledge is higher than that for physical capital (Ariel Pakes and Mark Schankerman, 1984).

The *q* equation imposes constant returns by assumption; I estimate only the relative implicit price of the *R&D* and capital stocks.

---

[19] This number should be taken with a large grain of salt. As noted above, the units of the pool variable are essentially arbitrary. The "true" relevant *R&D* for each firm could be 10 times smaller; this would not affect the elasticity estimates, but would lead us to conclude that Spence's θ is .3.

For firms with the average pool, the market apparently places more than three times as much value on a dollar of *R&D* stock as on a dollar of capital stock. Given the average returns implied by the profit equation, this is quite puzzling. It is possible that *R&D* spending conveys some sort of signal about long-run returns over and above its direct contribution to measured profits. It is also possible that this is an econometric artifact of limiting the sample to firms that report *R&D*.

Turning to the spillover effects, a different pattern is found than in the patent equation. The direct effect of the pool is to lower profits and market value. There is, however, an interaction between own and others *R&D* that implies that the return to own *R&D* is increased by the spillovers. For the firm with mean log(*R&D*), the *net* elasticity of profits with respect to the pool is about plus .1. If everyone increased their *R&D* by 10 percent, the aggregate profits would increase by about 3 percent, with about one-third of the net increase coming from the spillovers. For firms whose log(*R&D*) is about .5 standard deviations below the mean, the net effect of the pool is zero; for those with less *R&D* it has a negative effect. The pool effect in the q equation is quite similar, though much more significant statistically. Evaluated at the average log(*R&D*) and log(capital), the net elasticity of value with respect to the pool is about .05. For firms whose *R&D*-capital ratio is less than about one-half the average the net effect on value of the pool is negative.

The "competitive" (i.e., negative) effect of the pool is not significantly different from zero in either equation at conventional significance levels. The data do, however, reject at the 1 percent level the hypothesis that the log(pool) coefficient is zero in *both* equations. It appears that both technological spillovers and competitive effects of others' *R&D* come into play when we consider the economic returns to the firm's research. Comparing firms whose neighbors do a lot of *R&D* to those whose neighbors do little, the former are characterized by lower profits and market value if they do little *R&D* themselves, but a higher return to doing *R&D*. For firms with the average *R&D* budget, the net effect of others' *R&D* is positive.

A brief comment is in order on the effect of market share and concentration on profits and market value. Table 5 shows that higher market shares are associated with higher profits and market value, but increasing the four-firm concentration lowers the profits and market value of the average firm. Results not reported here show that each of these effects remains if the other variable is dropped, and interactions do not appear to be important. These results are consistent with other studies finding negative effects of concentration on profits at the firm or line-of-business level (Harold Demsetz, 1974; David Ravenscraft, 1983). They imply that the larger firms in a given industry have systematic advantages over their smaller rivals. It is not possible, with these aggregate firm data, to determine the extent to which the advantages are on the cost side, and the extent to which large market shares confer or reflect market power.

Finally, it is instructive to compare the *3SLS* results to the *OLS* results in Table 3. As already noted, the omitted variable story suggests that the coefficients on *R&D*, capital, and market share should be smaller in the *3SLS* results than in the *OLS*, since the former are consistent and the latter are biased upwards. This is clearly the case for market share. For *R&D*, however, the *3SLS* coefficients are consistently larger. This indicates that measurement error in *R&D* is a more serious problem than the correlation of *R&D* with left-out attributes.[20]

For completeness, the Appendix Table presents separate estimates for each year; the story they tell is essentially the same. I also estimated the equations separately (*2SLS*). A Hausman-type test comparing the *2SLS* results (not reported) to the *un-*

---

[20] Another aspect of these data support this argument. When I also estimate an *R&D* demand equation (not reported), I find that the residuals from that equation are negatively correlated with those from the contemporaneous patent, profit, and value equations. That is what we would expect from measurement error, but the omitted attributes story implies positive correlations.

constrained 3SLS results constitutes a specification test, since *2SLS* and *3SLS* are both consistent and hence should not differ much if the model is correctly specified. The *chi*-squared statistic for this test is 6.41 (12 degrees of freedom), which does not come close to rejecting the null hypothesis of correct specification.[21]

Classical *chi*-squared tests comparing the constrainted and unconstrained *3SLS* also fail to reject the null, which in this case is equality of the coefficients over time. The test statistics are .63 (3 degrees of freedom) for the patent equation, 3.8 (5 *d.f.*) for the profit equation and 3.2 (4 *d.f.*) for the q equation.[22] I also considered constraining the technological cluster dummy coefficients across time; these constraints were rejected for the profit and q equations;[23] none are imposed in the results of Table 5.

IV. Analysis of Technological Opportunity Effects

The coefficients of the technological cluster dummies in the profit, q, and patent equations measure, in some sense, the average excess returns (positive or negative) experienced by the firms in the group. In the patent equation, it is the average excess return to *R&D*, in terms of patents. In the profit and q equations, it is the average excess profits or market value remaining after controlling for the average rate of return on all of the firms' observed assets. These include *R&D*, capital, the spillover pool, market share, and industry concentration. The motivation for including these dummies in all the equations has been to allow for variations in technological opportunity, defined above as exogenous, technologically determined variations in the productivity of *R&D*. The *chi*-square statistics reported in Table 3 show clearly that the cluster dummies are important in all equations. It is difficult to know, however, whether technological opportunity is really the driving force behind the cluster effects. One way to investigate this question is to analyze and compare the patterns of coefficients across equations.

I have calculated the pair-wise correlations across the profit, q, and patent equations for the 21 dummy coefficients. The correlation across time is .62 for the patent equation, .26 for the profit equation, and .05 for the q equation. Within cross sections, the profit-q correlation is high (.84 for 1973; .80 for 1979), while the patent equation intercepts are relatively uncorrelated with those from the profit and q equations (.15 and .23 for 1973; .14 and .01 for 1979). It appears that the profit and q dummy coefficients share a common influence that is relatively fleeting; the patent coefficients are driven by something else that is relatively stable over time.

There are two possible interpretations of these results, one sympathetic to the technological opportunity story and one not. The sympathetic interpretation is that profits, q, and patents all respond to technological opportunity, but that, in the case of patents, this response is drowned out by the pattern of the propensity to patent, which is stable over time. The other view is that the absence of any correlation between the effects on patents and profit measures suggests that the cluster effects in the profit and q equations come from somewhere else. Implicit in the model as formulated is an assumption that all relevant influences from the market or demand side are captured by the industry variables that are in the model. To the extent that this is not true, then the cluster dummies could be capturing industry effects, since there is some correspondence between the clusters and industry groups.

Some additional insight can be gained by comparing these results to analogous ones where I substitute dummies based on industry groups for those based on technological clusters. Relevant data on the industry distribution of firms' sales are available only for 1972, so I can look only at the contemporaneous correlations. I find that the profit-

---

[21] The expected value of a *chi*-square with 12 degrees of freedom is 11.3. Note that the Hausman test often has very low power in this type of application, where both sets of estimators are inconsistent under the alternative hypothesis.

[22] Expected values are 2.4, 4.4, and 3.4, respectively.

[23] *Chi*-squares of 16.2, 75.4, and 83.1 for patents, profits, and value, respectively, compared to a 99.5 percent critical value of 40.

q correlation is .88, while the patent intercepts are correlated .09 with those from the profit equation and .07 with those with the q equation. Thus, when looking at industries, excess patents are even less correlated with excess profits and q, which are still highly correlated with each other. It seems that the data cannot really distinguish between industry effects and technological effects, though there is a small piece of the variation in profits that is associated with technology and not industries.

Regardless of whether excess profits or value are associated with industries or technological positions, it is interesting that they do *not* appear to be persistent over time; the correlations between the profit and q intercepts in each cross section are much larger than the correlation across time for either. This is consistent with the absence of significant mobility barriers, allowing firms to move gradually toward high profit clusters. Such movements would tend to compete away the excess returns. We would expect that, over time, the economic return to *R&D* in a "hot" area would fall, even if the potential of the area in a strictly technological sense remained high.

The lack of correlation across time could also result from the dependence of profits on exogenous factors which are not constant. There is, however, some evidence in the data that movement of firms played a role. As discussed above, I take the view that firms cannot change their position in technology space overnight, but that over some horizon they do move in response to perceived profit possibilities. An obvious test of this view is to ask if the movements in the firms' patent distributions *between* the 1965–72 period and the 1973–79 period bear any relation to the pattern of excess returns.

I calculated the proportional growth rate $((N_{79}-N_{73})/N_{73})$ of the number of firms in each technological cluster, and compared these to the cluster dummy coefficient in the 1973 q equation. There is a clear positive relationship. The correlation coefficient for the 21 observations is .53, which is significantly positive at the 1 percent level. The correlation of the growth rate with the dummy coefficients from the 1973 profit equation is .35. It appears that entry and exit were driven by the contemporaneous profit opportunities.

One would expect that this entry and exit would have an effect on the future realized profitability of the clusters. Again, the results seem to be consistent with this prediction. If we compare the *change* in the q equation dummy coefficients between the two cross sections with the growth rate of each cluster, the correlation coefficient is $-.79$, which is again highly significant. Even the 1979 intercepts themselves are negatively correlated with the growth rate, though somewhat less so $(-.56)$. Thus there is a clear association between entry (exit) into a cluster and a decrease (increase) in the average returns.

In interpreting these results, it is important to keep in mind that we are looking at snapshots, while the true process takes place in continuous time. It might appear as if the firms do not have rational expectations, because they choose to move toward what will be, on average, low-profit clusters. Such moves may well be rational, however, because the firms may make sufficient profits during the transition to make the move worthwhile—and they may well be in the process of moving on again to greener pastures.

One interesting question raised by these results is whether the "movers" are fundamentally different as a group from the firms that stay put. The above results regarding the effects of entry could be explained without reference to competitive forces if the firms that move are, on average, low-profit firms. To see if this is so, I examined the mean residuals for the two groups from the profit and q equations. The two equations for the two cross sections yield four comparisons of movers and nonmovers. In every case, the movers are actually *higher*-profit firms (though in three out of the four comparisons one cannot reject the hypothesis that the groups have the same mean). Thus, instead of weakening the entry story, this effect strengthens it. Entry is associated with a fall in the average returns for the group despite the fact that entrants are above average profit firms.

VI. Conclusion

I have presented an analysis of the effect on the productivity of *R&D* of certain factors from the supply side of innovation. In particular I find circumstantial evidence of spillovers of *R&D* from several indicators of technological success. Firms whose research is in areas where there is much research by other firms have, on average, more patents per dollar of *R&D*, and a higher return to *R&D* in terms of accounting profits or market value, though firms with very low own *R&D* suffer *lower* profits and market value if their neighbors are *R&D* intensive. All of these effects remain after controlling for the possibility that the technological areas themselves are associated with variations in the productivity of *R&D*.

Because the evidence is circumstantial, each piece taken alone is subject to interpretations not involving spillovers. For example, I had to assume that the spillover pool does not affect the propensity to patent. The result that I get (firms whose neighbors do a lot of *R&D* tend to get more patents after controlling for their own *R&D*) could, however, be due to firms' strategic decisions to patent more of their results if they perceive more technological competitors. When, however, one views the results from the three equations together, the combination of technological spillovers and competitive effects is a compelling interpretation.

I view these results as substantiating the spillover phenomenon. Understanding the implications for industrial structure and public policy will require more detailed work that focuses on the variations in the extent of spillovers in different areas, as well as examining the spillovers in a context where their implications for firms' strategic behavior are recognized.

The results regarding technological opportunity are less satisfactory. The patent class distribution data contain useful information about the composition of firms' research activities. Further, I find clear evidence that firms adjust their technological positions in response to profit possibilities, and that these adjustments tend to dissipate excess returns. It is not yet clear, however, to what extent the exogenous variations in profit potentials can really be attributed to technological opportunity rather than demand-side effects. Further examination of the "movements" of firms in technology space, combined with similar analysis of changes in firms' market positions, should provide more complete understanding of the interrelations among technological opportunity, changes in the structure of demand, profitability, and firms' actions in positioning their research and their sales.

APPENDIX

DATA SOURCES AND MANIPULATIONS

The data used herein come from four sources. First, the data regarding patent counts and their distribution by patent class come from the U.S. Patent Office, Office of Technology Assessment and Forecasting (OTAF). Second, firms' market share, their industry sales distribution, and the industry variables except *R&D* are from the Harvard Business School Program in Industry and Competitive Analysis (PICA) data base (Marilyn Shesko, 1981). These data were collected by PICA from two sources: the industry data are from the Census of Manufactures; the market share and sales distribution data are from a proprietary service and are derived from Bureau of the Census establishment data. The third data source was the FTC line of business report (U.S. Federal Trade Commission, 1979) for industry *R&D* numbers. All other data are from Standard & Poor's Compustat Tapes.

There are about 3000 nonduplicate manufacturing firms on the Compustat tapes—2600 of these (those that existed in 1976) were matched to the Patent Office data. Approximately 1700 firms had 1 or more patents. Of these, about 1200 reported positive *R&D* expenditure in 1976 (compared to 1500 of the 2600). Of the 1200, about 750 are available on the PICA data base. It should be noted that the firms lost going from the 2600 to the 750 are predominantly small firms.

These 750 firms are the universe from which a study requiring *R&D*, patent, and market-share data can be drawn. In order to characterize the firms' technological position using patent class distributions, a minimum of 10 patents was arbitrarily chosen. In each of the two periods (1965–72 and 1973–79), about 600 of the 750 meet this minimum—557 firms meet it in both periods; these are the firms that were used in the clustering analysis described in Table 1. For the regression analysis, complete financial data were needed for 1972–74 and 1978–80. This reduced the sample to 432 firms. Most of the regressions analysis described herein was also carried out for the 539 firms with complete data for the 1973 cross section only; the results are substantially similar (see my dissertation).

All flow variables are averages of three years data for each cross section. Stock variables are evaluated at the end of the middle year. The profit measure is operating income before depreciation *plus* reported *R&D* expense. The capital stock is the sum of plant and equipment, inventories, and investments in unconsolidated subsidiaries, each adjusted for inflation. Plant is adjusted for inflation by estimating the average age of the plant on the basis of the ratio of accumulated to current depreciation. Book value of plant is then multiplied by the ratio of the current *GNP* investment deflator to the deflator *n* years ago, where *n* is the estimated average age. Market value is the year-end value of common stock plus estimates of the value of preferred stock and bonds calculated by dividing dividend and interest payments by average market rates of return on these instruments. For more detail on both these variables, see Cummins et al.

The profit and value equation derivations refer to the stock of accumulated *R&D*. This is not something reported by the firm. Stocks must be constructed from reported *R&D* flows, making assumptions about the rate of obsolescence and *R&D* investments before the firms began reporting the annual *R&D* spending. Griliches and Jacques Mairesse (1984) experimented with several formulations; they found that results were relatively insensitive to the rate of obsolescence. They chose 15 percent per year, and that was utilized here as well. The stock at the end of the year before the first reported annual *R&D* is estimated by assuming constant exponential growth of annual *R&D*. Thus the initial stock is taken to be the reciprocal of the sum of the firm's *R&D* growth rate and 15 percent.

Four spillover pool variables were constructed, because the patent equation depends on the annual *R&D* flow while the other equations relate to the stocks. A pool of stocks and a pool of flows were constructed for each cross section using the weighted summation procedure described in Section I. Those for 1973 use as weights proximities based on the first-period distribution vectors, while those for 1979 use proximities based on the second-period distributions. For estimation, I took the log of these variables and subtracted off the mean, so that the *R&D* coefficients are *R&D* elasticities for the average-pool firm.

Finally, firm values for industry variables (*R&D*, shipments, concentration, growth rate, and minimum efficient scale) and market share were calculated as weighted averages of 4-digit SIC data, except for *R&D* which is only available for 193 industries. The weights are the industry fraction of the companies' 1972 sales, taken from the PICA data base.

APPENDIX TABLE—UNCONSTRAINED COEFFICIENT ESTIMATES
(432 Observations)

	log(Patents)		log(Profits)		log(Tobin's q)	
	1973	1979	1973	1979	1973	1979
log(*R&D*)	.881	.871	.109	.214		
	(.201)	(.195)	(.98)	(.062)		
R&D/Capital					3.84	3.12
					(.243)	(.155)
log(*R&D*)×log(Pool)	.321	.360	.024	.081		
	(.110)	(.061)	(.035)	(.044)		
(*R&D*/Capital) ×log(Pool)					.747	.818
					(.116)	(.092)
log(Pool)	.442	.521	−.072	−.091	−.011	−.064
	(.101)	(.121)	(.097)	(.044)	(.092)	(.022)
log(Capital)			.870	.899		
			(.062)	(.068)		
log(72 Share)			.187	.057	.305	.124
			(.052)	(.050)	(.061)	(.046)
log(Four-Firm Concentration Ratio)			−.240	−.206	−.549	−.520
			(.039)	(.030)	(.091)	(.118)
$\hat{\sigma}$	.832	.894	3.54	3.28	.609	.407

Note: For estimation details, see note to Table 5.

REFERENCES

Amemiya, Takeshi, "Non-linear Regression Models," in Zvi Griliches and Michael Intrilligator, eds., *Handbook of Econometrics*, Amsterdam, North-Holland, 1983.

Bernstein, Jeffrey I. and Nadiri, M. Ishaq, "Research & Development, Spillovers and Adjustments Costs: An Application of Dynamic Duality at the Firm Level," mimeo., New York University, 1983.

Bound, John et al., "Who Does R&D and Who Patents?," in Zvi Griliches, ed., *R&D, Patents and Productivity*, Chicago: University of Chicago Press, 1984.

Cummins, Clint et al., "The R&D Master File Documentation," mimeo., NBER, 1984.

Demsetz, Harold, "Two Systems of Belief About Monopoly," in Harvey Goldschmid, ed., *Industrial Concentration: The New Learning*, Boston: Little, Brown and Co., 1974.

Fisher, Franklin M. and McGowan, John J., "On the Misuse of Accounting Rates of Return to Infer Monopoly Profits," *American Economic Review*, March 1983, *73*, 82–97.

Griliches, Zvi, "Issues in Assessing the Contribution of R&D to Productivity Growth," *Bell Journal of Economics*, Spring 1979, *10*, 92–116.

______, "Market Value, R&D and Patents," *Economics Letters*, 1981, *7*, 183–87.

______, "*R&D, Patents and Productivity*, Chicago: University of Chicago Press, 1984.

______ **and Mairesse, Jacques,** "Productivity and R&D at the Firm Level," in Zvi Griliches, ed., *R&D, Patents and Productivity*, Chicago: University of Chicago Press, 1984.

Hall, Bronwyn, Griliches, Zvi and Hausman, Jerry, "Patents and R&D: Searching for a Lag Structure," NBER Working Paper No. 1227, 1983.

Hartigan, J. A., *Clustering Algorithms*, New York: Wiley & Sons, 1975.

Hausman, Jerry A., "Specification Tests in Econometrics," *Econometrica*, November 1978, *46*, 1251–72.

Hirschey, Mark, "Inventive Output, Profitability and Economic Performance," Graduate School of Business Working Paper No. 5-82-28, University of Wisconsin, 1982.

______ **and Wichern, Dean,** "Accounting and Market Value Measures of Profitability: Consistency, Determinants and Uses" *Journal of Business Economics and Statistics*, October 1984, *4*, 375–83.

Jaffe, Adam B., "Quantifying the Effects of Technological Opportunity and Research Spillovers in Industrial Innovation," unpublished doctoral dissertation, Harvard University, 1985.

Lindenburg, Eric and Ross, Stephen, "Tobin's *q* Ratio and Industrial Organization," *Journal of Business*, January 1981, *54*, 1–32.

McQueen, J. B., "Some Methods for Classification and Analysis of Multivariate Observations," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967.

Mansfield, Edwin, *The Production and Application of New Industrial Technology*, New York: W. W. Norton, 1977.

Nelson, Richard R., *The Rate and Direction of Inventive Activity*, Princeton: Princeton University Press, 1962.

Pakes, Ariel, "On Patents, R&D and the Stock Market Rate of Return," *Journal of Political Economy*, April 1985, *93*, 390–409.

______ **and Griliches, Zvi,** "Patents and R&D at the Firm Level: A First Look," in Zvi Griliches, ed., *R&D, Patents and Productivity*, Chicago: University of Chicago Press, 1984.

______ **and Schankerman, Mark,** "The Rate of Obsolescence of Patents, Research Gestation Lags, and the Private Rate of Return to Research Resources," in Zvi Griliches, ed., *R&D, Patents and Productivity*, Chicago: University of Chicago Press, 1984.

Ravenscraft, David, "Structure-Profit Relationships and the Line of Business and Industry Level," *Review of Economics and Statistics*, February 1983, *45*, 22–31.

Salinger, Michael A., "Tobin's *q*, Unionization, and the Concentration-Profits Relationship," *Rand Journal of Economics*, Summer 1984, 15, 159–70.

Scherer, F. M., "Firm Size, Market Structure, Opportunity and the Output of Patented Inventions," *American Economic Review*, December 1965, *55*, 1097–123.

______, "Demand Pull and Technological Innovation: Schmookler Revisited," *Journal of Industrial Economics*, November 1982, *30*, 225–37.

Schmookler, Jacob, *Invention and Economic Growth*, Cambridge: Harvard University Press, 1966.

Shesko, Marilyn, "Program for Industry and Competitive Analysis," mimeo., Harvard Business School, 1981.

Siu, Alan, "Essays in the Econometrics of Research and Development," unpublished doctoral dissertation, Harvard University, 1984.

Spence, A. Michael, "Cost Reduction, Competition, and Industry Performance," *Econometrica*, January 1984, *52*, 101–21.

White, Halbert, "Instrumental Variables Regression with Independent Observations," *Econometrica*, March 1982, *50*, 483–99.

National Science Foundation, *Research and Development in Industry, 1977*, Surveys of Science Resources Series, No. 79-313, Washington, 1979.

Explaining the Volume of Trade: An Eclectic Approach

By James R. Markusen*

Divide the world into a capital-abundant North and a labor-abundant South. Subdivide the North into two identical regions, East and West. Modern trade theory predicts intra-industry trade in differentiated manufactured goods between East and West, and traditional interindustry trade based on factor endowments between North and South. The model I develop explains both the volume and the direction of trade by combining nonhomothetic preferences with scale economies and differences in endowments.

Recent advances in international trade theory have increased our understanding of the causes and direction of international trade. While the factor-endowment model continues to play a prominent role in our theory, it is now widely accepted that differences in relative factor endowments are not the only cause of trade, particularly trade in manufactured products and presumably services.

An early approach to trade in manufactured goods was that of Staffan Linder (1961), who emphasized the role of demand. Linder noted that consumers with similar per capita incomes consume similar bundles of goods. Linder also viewed firms in the manufacturing sector as creating new products in response to perceived domestic demand. In addition to serving domestic customers, firms would eventually export to countries with similar per capita incomes since those countries would exhibit similar demands. Individual firms thus create products based on demand, and there will exist two-way trade in manufactured goods between two similar countries as each firm sells its good in both markets. The fact that both countries might have identical capital-labor ratios now actually encourages trade in the category of manufactures.

It was not until much later that the idea of two-way trade in similar manufactured goods between similar countries was formulated in consistent general-equilibrium models (Paul Krugman, 1979; Elhanan Helpman, 1981; Kelvin Lancaster, 1980; and Wilfred Ethier, 1982). These authors added scale economies, imperfect competition, and product differentiation in order to generate the type of trade suggested by Linder.

What we now have is a two-part theory of international trade. Divide the world into two regions, North and South, with equal total incomes but with the North being relatively capital abundant and the South relatively labor abundant. Now subdivide the North into two identical regions called East and West. ("East" should be thought of throughout the paper as a subset of the Western industrialized nations, not the communist block.) My theory predicts that N–S trade will be primarily the traditional Heckscher-Ohlin trade based on differences in factor endowments. It has become popular to refer to this as interindustry trade. E–W trade will be primarily trade in differentiated manufactured goods and services, referred to as intra-industry trade.

Although this theory combining factor endowments with scale economies gives us a good theory of the direction of trade, it has little to say about the volume of E–W vs. N–S trade. This point has indeed been made by Helpman (1981). One reason that the volume of trade is of interest is simply that E–W trade (i.e., trade among the industrialized countries) is very large relative to the volume of N–S trade.[1] Higher N–S trade

---

*Department of Economics, University of Western Ontario, London, Canada N6A 5C2.

[1] For an up-to-date survey and critical examination of the literature, see Alan Deardorff (1984).

barriers and a low total *GNP* in the South are part but probably not all of the story.[2] Linder's original empirical analysis did indeed correct for total *GNP*, but not for trade barriers. A second weakness of the theory is that it gives no role to demand in general and per capita income in particular in determining trade flows. Yet Linder's theory hypothesizes, and more recent empirical analyses show, that expenditure shares on commodities do depend on per capita incomes.

The purpose of this paper is to construct a simple model which provides a role for per capita income and is consequently able to analyze the total volume as well as the direction of E–N trade relative to N–S trade. The model is testable, and preliminary empirical results are very encouraging (discussed below). The key element is to combine monopolistic competition and differing relative factor endowments with a third determinant of trade, nonhomothetic demand. Interestingly, the model will put back into the theory one suggestion by Linder that was omitted by Krugman, Helpman, Lancaster, and Ethier. This is Linder's important observation that people with similar per capita incomes consume similar bundles of goods. I model this as an assertion that income elasticities of demand are quite different for different goods and therefore that utility functions are not homothetic. Other elements of Linder's theory such as the creation of new products and the role of sufficient domestic demand are not captured in the model. I am not claiming to model Linder's theory, only crediting him for suggesting an important role for per capita income.

The model developed below involves intraindustry trade of capital-intensive manufactured goods for one another and interindustry trade of a homogeneous labor-intensive commodity for manufactures. If we assume that the income elasticity of demand for manufactures exceeds one, while that for the homogeneous commodity is less than one, several interesting results follow directly. The labor abundant South is relatively specialized in producing the labor-intensive commodity, but also relatively "specialized" in consuming that commodity due to its lower per capita income. As the degree of nonhomogeneity increases, it is shown that the volume of E–W trade rises and the volume of N–S trade falls. This allows us to produce a Linder-type trading pattern based on a Linder-type demand assumption. But it clearly depends on the (empirically unestablished) connection between capital intensity in production and high income elasticity of demand in consumption. Independent of this assumption, the model does permit a clear and simple analysis of the role of demand in determining trade flows.

A few comments on the empirical relevance of the model are in order before proceeding. First, nonhomotheticity is certainly strongly supported by extensive empirical evidence on demand (a good summary is found in Angus Deaton and John Muellbauer, 1980). Formal work by Linda Hunter and myself (1985) has been done on the demand side of the model, estimating a linear expenditure system of the type developed below for a 34-country, 11-commodity 1975 data set using the real income data of Irving Kravis, Alan Heston, and Robert Summers (1982). The estimated deviations from homotheticity are highly significant in both statistical and economic terms. Food, in particular, has a very low estimated elasticity of demand and absorbs a major share of consumption expenditure in low-income countries. Expenditures on food ranged from 60 percent of personal consumption expenditures in India (lowest per capita income) to 13 percent in the United States (highest per capita income). Figure 1 plots the estimated share of food expenditure as a function of per capita income for the 34-country sample in Hunter's and my paper. Homotheticity would of course imply a constant share.

---

[2] Helpman's model (1984) does predict a large volume of E–W trade relative to N–S if (*i*) East and West have large *GNP*s relative to the South *and* (*ii*) differentiated manufactured goods constitute a large share of world *GNP*. Helpman seems to infer that (*ii*) is true given the observed trade volumes and the relative wealth of North vs. South. But this is hardly a test of any theory, as he points out. It is not at all clear to me how one would test this hypothesis since it requires that all goods be assigned to either the increasing-returns/differentiated or constant-returns categories.

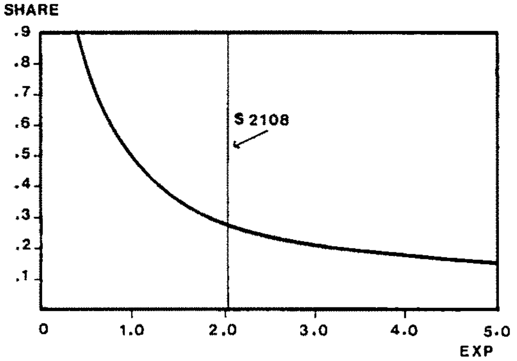


FIGURE 1. FITTED SHARE OF FOOD PLOTTED AGAINST TOTAL REAL PERSONAL CONSUMPTION EXPENDITURES PER CAPITA (Thousand of U.S. Dollars; Mean = $2108)

Source: Data: Kravis, Heston, and Summers (1982); Estimation: Hunter-Markusen (1986).

There is no formal analysis at this time to connect capital intensity in production with high income elasticity of demand in consumption, the combination of assumptions that generates a Linder-type concentration of trade among the industrialized countries. But "eyeballing" of the above-mentioned data set does suggest that there clearly exists a correlation between specialization in production and specialization in consumption for the top and bottom thirds of the countries. The low-income countries tend to specialize in both the production and consumption of food, while the high-income countries tend to specialize in both the production and consumption of manufactures and services (many of which are traded). As modeled here, it is these latter categories that tend to give rise to monopolistic-competition-type intra-industry trade. A formal analysis will be the subject of future work.

I. Demand and Production

Suppose that all individuals have identical utility functions given by

$$(1)\quad U=\left[(\Sigma x_i^{\alpha})^{1/\alpha}\right]^{\beta}(y-\gamma)^{1-\beta},\qquad 0<\alpha,\beta<1,$$

where y is a homogeneous commodity, and γ is a minimum consumption requirement. The x_i's are differentiated manufactured goods which are symmetric substitutes in consumption. y is used as numeraire, I_j is consumer j's income in terms of y, and p_i is the price of x_i. The assumption that y and not x has a minimum consumption requirement is arbitrary, but does have major implications as will become clear below.

Because the subutility function for manufactured goods is homogeneous of degree one, I can use a two-stage budgeting procedure in which the consumer first divides his income between manufactured goods and the homogeneous commodity. The second-stage problem involves choosing the various goods. In order to solve the first-stage problem, I define price and quantity indices for manufactures:

$$(2)\quad \tilde{x}\equiv(\Sigma x_i^{\alpha})^{1/\alpha};$$

$$\tilde{p}\equiv\min_{x}\left\{\Sigma p_i x_i\colon(\Sigma x_i^{\alpha})^{1/\alpha}=1\right\}.$$

The quantity index is thus just the subutility function for manufactures, while the price index is the minimum expenditure needed to buy one unit of the composite, or, alternatively, the expenditure function $e(p,\tilde{x}=1)$.[3] The consumer j's first-stage problem is then given by

$$(3)\quad \operatorname{Max} U_j^{*}=(\tilde{x})^{\beta}(y-\gamma)^{1-\beta}$$

$$\text{subject to}\quad (I_j-\tilde{p}\tilde{x}-y)=0.$$

The programming problem in (3) yields the following demand functions:

$$(4)\quad \tilde{x}=\beta(I_j-\gamma)/\tilde{p};$$

$$y=\gamma+(1-\beta)(I_j-\gamma).$$

The consumer's preferences and income consumption path are illustrated in Figure 2 for a particular price ratio $\tilde{p}$. The utility

---

[3] See Deaton and Muellbauer for a detailed analysis of this technique.

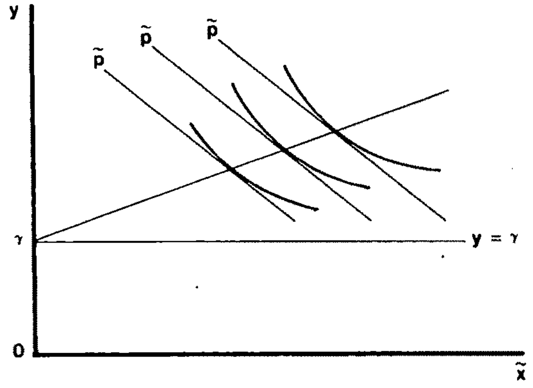


FIGURE 2

function is simply a Cobb-Douglas with its origin transposed to $(y=\gamma, x=0)$. The income consumption curve is linear, but does not go through the origin, indicating that the income elasticity of demand for x is greater than one, while that for y is less than one.

Because of these linear Engle's curves, I can aggregate individual demand functions in (4) to arrive at market demand functions, and I can indeed do so for the whole world if prices are the same everywhere. I shall make this free-trade assumption throughout this paper. Let X_i and Y denote total world demands for goods, L denote total world labor supply, and I total world income. Summing over individuals in (4), we obtain

$$\tilde{X}_i=\beta(I-\gamma L)/\tilde{p}_i; \qquad (5)$$

$$Y=\gamma L+(1-\beta)(I-\gamma L).$$

Note especially that market demands depend only on total income and total labor, not on the distribution of income among the individuals. The second-stage budgeting problem can be written as

$$\operatorname{Max} \tilde{U}_j=\left(\Sigma x_i^{\alpha}\right)^{1/\alpha} \qquad (6)$$

subject to $\left(\tilde{I}_j-\Sigma p_i x_i\right)=0;\ \tilde{I}\equiv\beta\left(I_j-\gamma\right).$

This programming problem yields familiar constant elasticity of substitution demand functions that are linear in $\tilde{I}_j$ and hence can be aggregated. Individual (x_{ij}) and total demands (X_i) for X_i are given by

$$x_{ij}=\tilde{I}_j/p_i^{\delta}\left(\Sigma p_i^{-\alpha\delta}\right); \qquad (7)$$

$$X_i=\tilde{I}/p_i^{\delta}\left(\Sigma p_i^{-\alpha\delta}\right); \qquad \delta=1/(1-\alpha),$$

where $\delta>1$ is the elasticity of substitution in consumption between X_i and X_j.

Turning to firms, we assume that Y is produced with constant returns to scale by competitive firms. Each firm in the X industry produces only a single product with increasing returns to scale. Products require a fixed cost $F(\omega)$ and a constant marginal cost $M(\omega)$, where ω is a vector of factor prices and where $F(\omega)$ and $M(\omega)$ are the same for all products (firms). A firm's average cost function is thus

$$AC_{x_i}=M(\omega)+F(\omega)/X_i. \qquad (8)$$

In order to make factor intensities unambiguous, it is assumed that the fixed and variable inputs use factors in the same relative proportions. Thus the total factor proportions used in X depend only on ω and not on the scale of output.

These identical cost functions plus the demand functions in (7) imply that it is always better to produce a different product than a product currently produced by an existing firm. Hence all products will be unique and each consumer will buy some of each product. Further, free entry by (single product) firms ensures that all X_i's that are produced sell for the same price and are produced in the same quantity. In what follows I shall therefore let p and X denote the price and quantity of a representative product and let n denote the number of products produced.

Each firm in the X industry is assumed to have a small market share and consequently I make the usual monopolistic competition assumption that firms view the marginal utility of income as fixed. In the present model, this amounts to assuming that $\tilde{I}/(\Sigma p_k^{-\alpha\delta})$ is viewed as constant in the demand functions (7).[4] The elasticity of demand as perceived by an individual producer is thus found by differentiating (7) holding $\tilde{I}/(\Sigma p_k^{-\alpha\delta})$ constant:

---

[4] This assumption was first used in trade theory by Krugman.

$$\eta \equiv (\partial X_i/\partial p_i)(p_i/X_i) = -\delta; \tag{9}$$

$$\text{or} \quad 1/\eta = -1/\delta = (\alpha - 1).$$

Marginal revenue for an individual X producer is then given by the usual formula

$$MR = p(1+1/\eta) = p\alpha, \tag{10}$$

which is of course less than p and decreasing in X. Monopolistic competition equilibrium is found from the equations $MR = MC$ and $p = AC$.

$$p\alpha = M(\omega) \tag{11}$$

$$p = M(\omega) + F(\omega)/X. \tag{12}$$

Equations (11) and (12) are solved to yield output per firm

$$X = \alpha F(\omega)/((1-\alpha)M(\omega)). \tag{13}$$

Note that the level of output of each variety does not depend on total income or on per capita income, but only on equilibrium factor prices.

From (5) above, we have

$$pnX = \tilde{p}\tilde{X} = \beta(I - \gamma L). \tag{14}$$

Replace X in (14) with (13) and replace p in (14) with $M(\omega)/\alpha$ from (11). This gives us the equilibrium number of varieties produced:

$$n = (\beta(I-\gamma L)(1-\alpha))/F(\omega). \tag{15}$$

Combining (14) and (15) we see that the X industry expands at constant factor prices by adding varieties rather than by increasing the outputs of individual goods.

II. The Integrated World Equilibrium

Consider the equilibrium that would emerge if we lumped all regions together or alternatively allowed both factors and goods to move costlessly between regions. Avinash Dixit and Victor Norman (1980) refer to this as the "integrated world equilibrium," a concept that has been extensively exploited by Helpman and Krugman. Let $\bar{p}$ and $\bar{\omega}$ be the equilibrium commodity price ratio and vector of factor prices, respectively, at the integrated world equilibrium. For the remainder of the paper assume also that there are only two factors of production, capital and labor (K and L), and that X is capital intensive. Finally, assume that the population of a country is identical to L and therefore that there is not a separate class of people called capitalists. Equivalently, ownership of capital is distributed (not necessarily evenly) among labor.

At a given set of factor prices, a country's per capita income will therefore be a function of its capital-labor endowment ratio. With $\gamma > 0$, the country with the higher K/L ratio will spend a larger fraction of its income on X. Note also that because my utility function permits perfect aggregation, it is not necessary to assume that workers within a country have identical capital endowments.

Suppose that in the integrated equilibrium the total amount of (K, L) required in the X sector is B in Figure 3, and that C gives the total requirements in the Y sector. Assume initially that $\gamma = 0$ and $\beta = 1/2$, so that the value of total production is the same in each sector ($pnX = Y$) and equal to ($I/2$). The

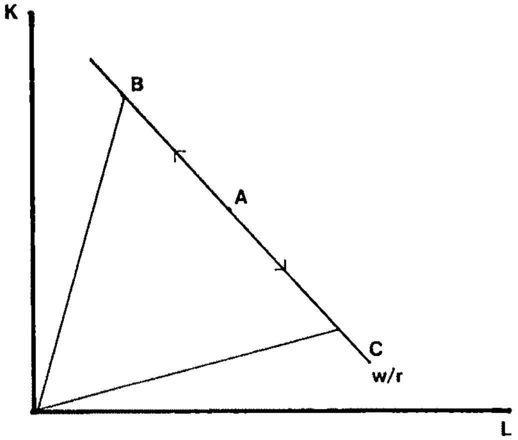


FIGURE 3

costs of factor bundles B and C are thus equal, and the slope of the line passing through B and C is (w/r).

In order to provide a meaningful comparison of different world equilibria, I shall in what follows consider "compensated" changes to the model that preserve the integrated world equilibrium. This allows me to compare the bilateral trade volumes at the same prices, total production quantities, and regional *GNP* levels.

III. Benchmark Equilibria

In this section I characterize two pure cases, both of which are well-known. In each case I assume homothetic demand $(\gamma = 0)$ and assume that $\beta = 1/2$. One is the case of pure intra-industry trade (Krugman) and the other is the case of pure interindustry trade, the Heckscher-Ohlin model. In order to limit the dimension of the taxonomy, I will assume that the North and South have identical absolute levels of *GNP* so as to avoid well-understood determinants of trade volumes arising from such differences.[5]

A. *Pure Intra-Industry Trade*

Consider first the case of pure intra-industry trade. I divide the world into two regions, N and S, and give each region exactly one-half the world's factor endowment. The endowments of N and S are thus given by A in Figure 3. Region N can be further subdivided into E and W, of identical size and relative factor endowments. This allocation of factors clearly has the integrated world equilibrium as its equilibrium and can be characterized as follows.

1) N and S are diversified, each producing $Y = I/4$ and differentiated products $pnX = I/4$. The variables p and X are the same everywhere so the South produces the same number of varieties (one-half the total) as N, or twice the number of E and W individually.

2) There is no trade in Y, and S exports 1/2 the output of each of its X_i varieties to N, which is 1/4 of S's *GNP*. This is because each individual, no matter where he (she) is located, consumes the same amount of every variety produced as does any other consumer. Subregion E sends 3/4 of the output of each of its varieties abroad, with 1/2 going to S and 1/4 going to W.

Pure Intra-Industry Trade: N *and* S *Identical*:

S exports (1/4) of its *GNP* to the North, or (1/8) of its *GNP* to E and W individually.

E exports (1/8) of its *GNP* to the West and (1/4) of its *GNP* to S; similarly for W.

All trade is intra-industry trade.

The share of trade in a region's *GNP* is proportional to one minus the share of that region's *GNP* in world *GNP*.

B. *Pure Interindustry Trade*

Now consider an equilibrium in which there is only interindustry trade. Starting at the integrated world equilibrium, let $\alpha = 1$ and $F(w) = 0$ so that both the taste for diversity and scale economies disappear. The X industry becomes perfectly competitive producing with constant returns. In order to generate an integrated equilibrium with the same p, I do however need to raise the marginal cost of production $M(\omega)$ to equal the old average cost. From (11), the old average cost is equal to $M(\omega)/\alpha$, so I can define a new marginal cost function as $M*(\omega) \equiv M(\omega)/\alpha$.[6]

Given $\alpha = 1$, $F(\omega) = 0$, and $M*(\omega) = M(\omega)/\alpha$, the integrated equilibrium will have the same p, ω, and I as just discussed. Since $\alpha = 1$, I only need the demand functions (5) from the first-stage budgeting problem to obtain all relevant demand information. The competitive price equals marginal cost condition $p = M*(\omega) = M(\omega)/\alpha$ similarity ensures the same supply relationship. If both N and S were at point A in Figure 3, there would be no trade or, more correctly, entirely arbitrary trade given the assumed absence of tariffs and transportation costs.

---

[5] See Helpman (1984) for a discussion of the country size (total *GNP*) issue.

[6] This procedure relies on my earlier assumption that factor proportions depend only on ω and not on the level of output.

Now slide the endowments of N and S apart along w/r in Figure 3, until N reaches point B and S reaches point C. Subregions E and W continue to be identical halves of N. Regions N and S will now specialize in X and Y, respectively, and the integrated equilibrium values of p, ω, and I will be preserved. Further, N and S will each have an income equal to $(I/2)$ and will wish to spend equal shares of income on X and Y. This Heckscher-Ohlin equilibrium can be summarized as follows.

Pure Interindustry Trade: N *and* S *Specialized,* $\alpha=1$, $F(\omega)=0$, *and* $M*(\omega)=M(\omega)/\alpha$:

S exports (1/2) of its *GNP* to the North or (1/4) of its *GNP* to E and W individually.

There is no E–W trade, and each of these regions exports (1/2) of its *GNP* to the South.

All trade is interindustry trade.

The share of trade in each region's *GNP* is the same and therefore unrelated to the region's size.

I should note that the last presumption above is a very special result holding only for the case chosen (it relies on the total *GNP* of N equalling that of S, and on E and W being identical). Nevertheless, it does show that in a multiregion world there can be no presumption that the share of trade in *GNP* falls with country size. It is interesting to note that this presumption does hold in the case of pure intra-industry trade, but not in the Heckscher-Ohlin case.[7]

IV. Mixed Trade Caused by Endowment Differences, Scale Economics, and Nonhomothetic Demand

Let us return α, $F(\omega)$, and $M(\omega)$ to their original values and consider again the integrated equilibrium with $\gamma=0$ and $\beta=1/2$. Beginning at point A in Figure 3, we can once again slide N and S apart to B and C, respectively. At the initial p, ω, and therefore I, both demand and cost conditions will be unchanged. Per capita income is higher in the North but, as noted in Section I, the demand function facing X producers depends only on total I. The consequence is that the integrated equilibrium values of p, ω, and I continue to be equilibrium values. The only difference between the B/C equilibrium and the A equilibrium in Figure 3 is that N and S are specialized in the former. Recall that each individual, no matter where he (or she) is located, consumes each variety of X and that expenditure shares on X and Y are still (1/2) in each region. Trade volumes are as follows.

Mixed Trade with Homogeneous Demand ($\gamma=0$):

S exports (1/2) of its *GNP* to the North or (1/4) of its *GNP* to the E and W individually.

E exports (1/4) of its *GNP* to the W and (1/2) of its *GNP* to the S.

N–S trade is exclusively interindustry and E–W trade is exclusively intra-industry.

The volume of E–W trade *relative* to N–S trade is not affected by the spread in endowments. Both volumes double relative to the case of pure intra-industry trade.

The volume of E–W trade increases relative to the case of pure interindustry trade while the volume of N–S trade is unaffected. Scale economies and product differentiation thus create intra-industry trade, but do not displace interindustry trade.

Now let us turn to the role of nonhomogeneous demand and consider first initial integrated equilibrium with $\gamma=0$ and $\beta=1/2$. Suppose that we increase both γ and β from these initial values in a manner such that $\beta(I-\gamma L)$ in equation (5) stays constant at $(I/2)$ for the initial value of I. As in previous examples, this "compensated" change will preserve the initial p, ω, and I as equilibrium values. From (5) it can be seen that total expenditure shares on X and Y will continue to be 1/2 at the initial I, and the demand functions in (7) will be unchanged at the initial I. Equations (13) and (15) similarly give the same values of X and n at the initial I and ω. The initial p, ω, and I are preserved and points B and C in

---

[7] Also noted by Helpman (1984).

Figure 3 continue to give the factor bundles needed to produce nX and Y.

As before, we can give N and S factor bundles B and C, respectively, in Figure 3. This does not affect the total demand for X or the elasticity of demand for X and so the equilibrium values of p, ω, and I are equal to their integrated world equilibrium values. There is a world expenditure share on X of (1/2), and total North and South incomes are unaffected and remain at $(I/2)$.

At point A in Figure 3, N and S have equal per capita incomes as well as total incomes. But as S slides toward C in Figure 3, L_s (South labor force) grows while $I_s = I/2$ remains constant. Let X_s^* denote the *value* of total South demand for all the X_i. The change in the value of South demand for X follows from (14):

$$dX_s^* = -\beta\gamma dL_s. \tag{16}$$

If L_s^a denotes the South's labor force at A and L_s^c its labor force at C, the fall in demand for X in the South is

$$\Delta X_s^* = -\beta\gamma(L^c - L^a). \tag{17}$$

Note from (14) above that a fall in demand for X by S is exactly matched by the rise in demand for X by N, so trade remains balanced and equilibrium is preserved as noted above.

Let $\sigma \equiv X_s^*/I$ be the share of expenditure on X. Then the South's share σ_s at C is

$$\sigma_s = \frac{\beta(I_s - \gamma L_s^a)}{I_s} - \frac{\beta\gamma(L_s^c - L_s^a)}{I_s} = \left(\frac{1}{2}\right) - \frac{\beta\gamma(L_s^c - L_s^a)}{I_s} < \frac{1}{2}. \tag{18}$$

As noted above, L_s^c, L_s^a, and I_s are all unaffected by the offsetting increases in β and γ described above. Thus it is possible to raise γ and β (such that $\beta(I - \gamma L^a)$ is constant) to the point where σ_s in (18) is arbitrarily close to zero.

Mixed Trade with Nonhomothetic Demand $(\gamma > 0)$:

S exports $(\sigma_s < 1/2)$ of its *GNP* to the North or $(\sigma_s/2)$ of its *GNP* to the East and West individually.

E exports $[(1 - \sigma_s)/2] > (1/4)$ of its *GNP* to the W and $(\sigma_s < 1/2)$ of its *GNP* to the South.

The volume of E–W trade increases relative to the volume of N–S trade as a consequence of the spread in endowments. The former more than doubles and the latter less than doubles relative to the case of pure intra-industry trade.

The volume of E–W trade increases relative to the case of pure interindustry trade while the volume of N–S trade falls. Nonhomotheticity of the type assumed here thus reduces the volume of interindustry trade. The volume of trade between countries with similar relative endowments can exceed that between countries with dissimilar relative endowments.

These results rely of course on the assumption that the labor-intensive good has the less-than-unitary demand elasticity. This assumption means that each country has a high propensity to consume its own export good and therefore that interindustry trade is reduced relative to the case of homothetic demand. But intra-industry trade among the similar capital-abundant countries (East and West) is simultaneously increased.

This result concerning the effect of nonhomothetic demand on the volume of interindustry trade is illustrated in Figure 4. Let TT' denote the identical production frontiers for N and S when they have the same endowments. This frontier depends on the output per firm in the X industry, which I assume equals the output at the integrated world equilibrium. The frontier is then convex to the origin, since expansion of X takes the form of adding firms rather than changing the output per firm as noted above. A is the identical consumption point for N and S and $\tilde{p}$ the equilibrium price ratio. Each country's community indifference curves are homothetic with respect to the displaced origin $(Y = \gamma L/2, X = 0)$.

Production frontiers SS' (for the South) and NN' (for the North) in Figure 4 correspond to the endowments C and B, respectively, in Figure 3. The South specializes in Y and the North in X. As discussed earlier, the integrated equilibrium values of p, ω, and I are preserved. If demand was homothetic, each country would continue to con-

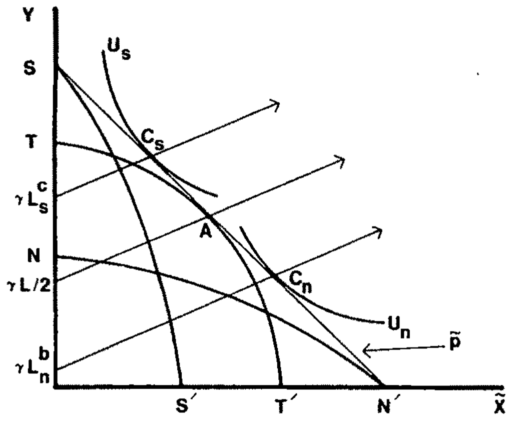


FIGURE 4

sume at A and the North and South would have trade vectors AN' and SA, respectively.

With nonhomothetic demand, the differing per capita incomes of N and S make demands different in those two regions. Demand for the South is homothetic with respect to the origin ($Y = \gamma L_s^c$, $X = 0$) and demand by the North with respect to ($Y = \gamma L_n^b$, $X = 0$). The integrated equilibrium values of p, ω, and I are preserved, but the South and the North consume at C_s and C_n, respectively. The volume of interindustry trade is now reduced relative to the case of homothetic demand, with N and S now trading C_nN' and SC_s, respectively.

If I were to reverse either the factor-intensity assumption or the income-elasticity assumption, the volume of trade would of course increase relative to a Heckscher-Ohlin (homothetic demand) world. In Figure 4, if Y were capital intensive, then country S would produce at N' and country N would produce at S. Their consumption bundles would continue to be as shown, indicating a very large volume of interindustry trade.

V. Summary and Conclusions

The purpose of this paper was to construct a model that could offer empirically plausible predictions about both the volume and direction of trade in a multiregion world. As is often done, the world here is decomposed into a capital-abundant North and a labor-abundant South. The North is then further divided into two identical regions, East and West. One production sector produces differentiated, capital-intensive manufactured goods with increasing returns to scale while the other sector produces a homogeneous, labor-intensive commodity with constant returns to scale.

The existing theory based on homothetic demand predicts that E–W trade will be intra-industry trade in differentiated manufactured goods while N–S trade will be inter-industry trade of the homogeneous good for manufactures. But while this theory gives us a good model of the direction of trade, it cannot adequately describe the volume of E–W vs. N–S trade. This was demonstrated here by first considering the case where N and S had identical absolute and relative capital-labor endowments (Section III). When I then transferred labor from N to S and capital from S to N, I showed that both E–W intra-industry trade and N–S interindustry trade increased in the same proportion. In the example of Section IV, both E–W and N–S trade doubled as a consequence of the spread in endowments.

The factor-endowments cum scale-economies model is thus not rich enough to offer insights into the volume of E–W vs. N–S trade. It cannot by itself offer explanations as to the large volume of E–W vs. N–S trade. There are, of course, several well-known explanations of this unequal trade volume. First, N–S trade barriers are thought to be significantly higher than E–W trade barriers. Second, the total GNP of the South is smaller than that in the North. Linder's original empirical analysis corrected for the latter but not for the former.

This paper takes a different approach to trade volume by following Linder in assuming nonhomothetic preferences. A Linder-type trading pattern (trade concentrated among the capital-rich industrialized countries) emerges if capital intensity in production is correlated with high-income elasticity of demand in consumption. The capital-rich North specializes in manufactured goods, but also spends a large fraction of its income on these goods due to its high per capita income. Conversely, the South spends a large fraction of its income on its own export good. It was demonstrated that as the degree of nonhomogeneity increases in a well-defined way, the absolute level of E–W trade increases while the absolute level of N–S trade falls. It must be emphasized that the empirical connection between factor intensities and income elasticities has not yet been established. This Linder-type result is thus only a special case of the paper's more general contribution: modeling how per capita income can systematically influence trading patterns.

To summarize, the model has three bases for trade: differences in relative factor endowments, scale economies and product differentiation, and nonhomothetic preferences. The first two are needed to explain the direction of trade and the third is needed to explain the volume of E–W vs. N–S trade. This third factor, although new to formal trade theory, is strongly supported by empirical evidence.

REFERENCES

Deardorff, Alan, "Testing Trade Theories and Predicting Trade Flows," in Ronald Jones and Peter Kenen, eds., *Handbook of International Economics*, Vol. 1 Amsterdam: North-Holland, 1984.

Deaton, Angus and Muellbauer, John, *Economics and Consumer Behavior*, Cambridge: Cambridge University Press, 1980.

Dixit, Avinash and Norman, Victor, *Theory of International Trade*, Cambridge: Cambridge University Press, 1980.

Ethier, Wilfred, "National and International Returns to Scale in the Modern Theory of International Trade," *American Economic Review*, June 1982, *72*, 389–405.

Helpman, Elhanan, "International Trade in the Presence of Product Differentiation, Economies of Scale, and Monopolistic Competition," *Journal of International Economics*, August 1981, *11*, 305–40.

______, "Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries," Working paper, Tel-Aviv University, 1984.

______ **and Krugman, Paul,** *Increasing Returns, Imperfect Competition, and Foreign Trade*, Cambridge: MIT Press, 1984.

Hunter, Linda and Markusen, James, "Per Capita Income as a Determinant of Trade," Working Paper, University of Western Ontario, 1986.

Kravis, Irving, Heston, Alan and Summers, Robert, *World Product and Income: International Comparisons of Real Gross Product*, World Bank, Baltimore: Johns Hopkins University Press, 1982.

Krugman, Paul, "Increasing Returns, Monopolistic Competition, and International Trade," *Journal of International Economics*, November 1979, *9*, 469–79.

Lancaster, Kelvin, "Intra-Industry Trade under Perfect Monopolistic Competition," *Journal of International Economics*, May 1980, *9*, 151–75.

Linder, Staffan, *An Essay on Trade and Transformation*, Uppsala: Almqvist and Wiksells, 1961.

Nominal Contracts in a Bimetallic Standard

By PETER M. GARBER*

A bimetallic standard grants nominal debtors an option to deliver either of two metals. Construction of theoretical bimetallic option values allows the computation of the option value implicit in nineteenth-century U.S. securities. Appropriate yield adjustments permit a comparison of U.S. and monometallic European security yields. Yields on antebellum U.S. securities are close to British yields. Evaluating the option also allows the estimation of wealth transfers from shifts in the monetary standards during the nineteenth-century silver agitation.

The objective of this paper is to determine the value of contracts to deliver dollars under a bimetallic standard, a multiple commodity standard of the sort embraced by many of the industrializing countries in the nineteenth century. Since contracting in bimetallic dollars grants the short interest an option to deliver either gold or silver, a dollar contract can be evaluated by applying results from the options literature. With such an evaluation, one can determine the magnitude of wealth transfers effected by terminating or returning to a bimetallic standard, as occurred during phases of the nineteenth-century silver agitation. Also, evaluation of the bimetallic contract provides a rationale for the high yields on dollar securities relative to sterling yields observed in the nineteenth century. While high dollar yields have been interpreted in the literature as an indicator of a higher cost of capital in the United States, they resulted merely from the difference in monetary standards.

Under a bimetallic monetary standard, the unit of account is defined as either a given weight of gold or a given weight of silver.[1] Since the market relative price between the two metals rarely equals the legally implied exchange ratio, the cheaper metallic unit or its value equivalent is delivered to satisfy contracts in the nominal unit.[2]

Most discussions of bimetallic standards focus on the necessary appearance of a sequence of monometallic epochs which alternate between deliveries of gold and silver money.[3] The observed failure of most bimetallic systems to achieve the simultaneous circulation and delivery of both metals is deemed obvious evidence of the weakness of bimetallism. Economists have proposed more sophisticated metallic or general commodity

---

*Department of Economics, Brown University, Providence, RI 02912. I am grateful to Cliff Smith, Stan Engerman, Neil Wallace, Tony Kuprianov, J. S. Butler, Paul Romer, Bob King, Lauren Feinstone, Naomi Lamoreaux, and two referees for useful comments. I have also benefited from comments by participants in seminars at Northwestern University, Brown University, Yale University, the Wharton School, North Carolina State University, MIT, and the Federal Reserve Bank of Minneapolis. Luis Suarez provided helpful and resourceful research assistance. Research for this paper was supported by NSF grant SES-8319627.

[1] More accurately, legal tender coins defined as equal to a certain number of units of account contain precisely defined weights of one or the other precious metal.

[2] Arthur Rolnick and Warren Weber (1986) have recently discussed circumstances under which Gresham's Law will not apply, using the U.S. bimetallic and greenback periods as examples. Even in such circumstances the value equivalent of the overvalued metal will be delivered to satisfy dollar contracts.

[3] This is particularly true of J. Laurence Laughlin (1885) who was strongly opposed to bimetallism. However, less partisan writers such as Alfred Marshall (1887), Irving Fisher (1923), and Robert Barro (1979) also raise this issue. For a concentration on the delivery option dimension of bimetallism, see William Jevons (1899) who characterized bimetallism as an "alternating standard." The delivery option feature was recognized as the essence of bimetallism by its supporters, who also clearly understood the necessary alternation of monometallic epochs under bimetallism. See U.S. Congress (1877, pp. 11, 91–101) for a clear statement of the nature of the option feature from the pro-bimetallic camp.

standards, such as symmetallism or tabular standards, aimed both at circumventing this monometallic circulation and at stabilizing the price level.[4] While many economists expect welfare gains from price level stabilization, it is not clear why recurrent switching from one circulating metal to another should be detrimental of itself.

Policy discussions of the U.S. bimetallic standard peaked in the 1890's and 1930's. Since these were periods of depression, deflation, and cheap silver, powerful coalitions of groups which would benefit from inflation organized around the reestablishment of the original bimetallic dollar standard. In such a system, cheap silver-based money would immediately become the circulating medium, driving up the price level and effecting substantial transfers to debtors.

This paper develops a method to price actual nominal contracts under bimetallism and to address some historical issues arising from the U.S. bimetallic experience. In the development of a pricing model and of the estimates of the bimetallic option values, I follow the same strategy as Gerald Gay and Steven Manaster (1984), who constructed the option values of delivering different qualities of wheat in a futures contract. Section I presents a brief history of nineteenth-century U.S. coinage laws relating to bimetallism. Section II contains a demonstration of how to evaluate a bimetallic contract using results from the options pricing literature. In an example based on results in Robert Merton (1973), William Margrabe (1978), and René Stulz (1982), the current value of a contract to deliver dollars in the future is developed for continuous time environments. One can readily show that the price paid for a dollar contract will be less than the prices paid for otherwise identical contracts promising to pay only gold or only silver. The measured yields on dollar contracts will then be higher than yields on monometallic contracts.

In Section III, I compute theoretical values of actual U.S. Treasury securities promising to deliver bimetallic dollars in varying future payment streams. I calculate the implied option value of the bimetallic bonds relative to gold bonds and compare bimetallic theoretical bond prices to actual realized prices. In Section IV, I address the issue of whether capital was more costly in the nineteenth-century United States than in Great Britain, a standard debate in the historical literature. In Section V, I use the computed option values to examine the potential wealth transfers from terminating the free coinage of silver in 1873 and restoring it at the end of the nineteenth century. The results indicate that the wealth transfer from debtors associated with the termination of the free coinage of silver in 1873 was small, amounting to no more than one year's interest payment on outstanding debt. Given the low price of silver in the 1890's, restoration of free coinage of silver would have far more than offset the value of the original transfer.

I. The Evolution of U.S. Bimetallism

Here I present a brief chronology of the legal development of U.S. bimetallism. Also, I describe some of the legislation implemented in the post-1873 agitation to restore the free coinage of silver. I have relied on J. Laurence Laughlin (1885), Charles Dunbar (1883), the U.S. Senate Finance Committee (1894), and Milton Friedman and Anna Schwartz (1963) as sources for this discussion.

Prior to the adoption of the Constitution, foreign coins, primarily silver Spanish dollars, comprised the specie of the United States. The first national coinage was provided under the Act of 1792 which established the mint and defined the dollar as 371.25 grains (.773 troy ounces) of silver. The law authorized the coinage of a silver dollar and half-dollars, quarter-dollars, dimes, and half-dimes containing corresponding weights of silver. In addition, the mint was authorized to issue ten-dollar gold coins (eagles) containing 247.5 grains (.515 troy ounces) of gold, as well as half-eagles and quarter-eagles. Effectively, the law defined the monetary value of a given weight of gold as fifteen times that of the same weight of silver. It provided for free coinage of both metals in that anyone delivering a given weight of a metal to the mint could receive in exchange authorized coins containing an equal weight of the same metal. This service would be provided free of charge if the delivering party were willing to allow the coinage to occur "as speedily as may be after the receipt thereof [of the bullion]" (Dunbar, p. 228). Finally, the law declared that the coins struck at the mint were lawful tender "in all payments whatsoever, those [coins] of full weight according to the respective values herein before declared" (Dunbar, p. 229).[5]

---

[4] Symmetallism prescribes the definition of the dollar in terms of a basket of metals, where a given weight of each metal is contained in the basket. There is no option associated with the short interest. In a tabular standard, the dollar is valued according to a particular price index. See, for example, Milton Friedman (1984) for a recent description of a tabular standard and Marshall (1887) on symmetallism.

Figure 1 indicates the ratio of the market price of gold to that of silver from 1787 to 1884. In 1792, as the coinage act was passed, the market ratio was approximately 15/1. However, the ratio permanently rose above 15/1, making silver overvalued at the mint. Since contracting parties would then deliver only silver or its value equivalent to satisfy dollar contracts, the real value of the dollar was the value in terms of goods of 371.25 grains of silver.

In an act of 1834, the dollar was devalued in terms of gold. The ten-dollar gold coin would now contain 232. grains (.483 troy ounces) of gold, a devaluation of 6.3 percent, which changed to 16/1 the weight ratio of silver to gold in the dollar. Proportional changes were made in the gold content of the smaller gold coins. Since the market relative price fluctuated between 15.5 and 16., this devaluation meant that gold or its value equivalent would now be delivered to satisfy dollar contracts. The act of 1837 slightly revalued the ten-dollar coin to 232.2 grains of gold, with corresponding changes in the smaller gold coins.

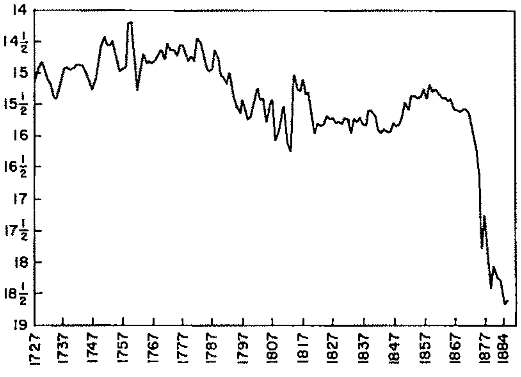


FIGURE 1. THE NUMBER OF OUNCES OF SILVER EXCHANGEABLE FOR ONE OUNCE OF GOLD

Source: Laughlin, p. 161.

In 1853, the full silver weight and unlimited legal tender properties of the fractional coinage were terminated. The amount of silver in the fractional coins was reduced by 7 percent, and they were made legal tender in quantities not exceeding five dollars. Furthermore, the free coinage of fractional coins was terminated. This change meant that the ratio of silver contained in a dollar's worth of fractional coins to gold in the gold dollar was less than 15/1. The law left unchanged the definition of the silver dollar.

In 1859, the enormous Nevada Comstock Lode was discovered, a silver strike which reversed the decline in the relative value of gold begun with the Californian and Australian gold strikes of the 1840's. Silver prices slowly declined until 1873 when the accelerating development of the Nevada deposits initiated a rapid fall in the silver price.

In the act of February 12, 1873, silver was effectively demonetized. The old silver dollar was not included among the silver coins authorized for minting in the United States, although a "trade dollar" and fractional coins were allowed. Only these authorized silver coins were given legal tender status for payments not exceeding five dollars. Free coinage of legal tender silver was eliminated. By this act the metallic dollar was effectively defined as a given weight of gold alone.[6] Any contracts for future delivery of dollars were redefined by this act in that only gold or greenbacks could be delivered, where before silver could have been delivered. From the point of view of the government, which delivered only specie on its debt, its contractual option to deliver silver was now removed.[7] In addition to the United States, other important countries abandoned silver simultaneously, that is, Germany which had a silver standard and the Latin Union (France, Italy, and Belgium) which had a bimetallic standard.

---

[5] Why a bimetallic standard was the preferred standard of the political process is not clear. Although carefully considered, the reasons for selecting the 1792 system outlined by Alexander Hamilton (1791) seem to have had their basis in the existing system of circulating coinage bequeathed to the United States.

Some researchers have argued that the changes had little practical effect since the market ratio of the gold price to the silver price was still below 16/1 in 1873.[8] However, though the 1873 market ratio was still below 16/1, this change may have represented an important wealth transfer. If the value of the bimetallic option to pay in silver was high, anyone engaged on the short side of a dollar contract prior to 1873 would immediately feel the impact of the change in standard as a transfer from himself to his creditors. Similarly, taxpayers would recognize a transfer from themselves to the government's creditors.

Soon after the law's passage, the market ratio exceeded 16/1, reaching levels over the next twenty years which were unprecedented in the prior two centuries of data. The sudden rise in the market ratio triggered a quarter-century agitation to restore the free coinage of silver. First fruit of this movement was the Bland-Allison Act of 1878 which reauthorized the issue of the standard silver dollar with legal tender status and required limited Treasury silver purchases with silver certificates. Since government obligations required only the delivery of coin, this change restored the bimetallic option to the government. However, free coinage of silver was not restored.

The Sherman Silver Purchase Act of 1890 repealed the silver purchase provisions of the Bland-Allison Act, replacing them with a requirement to purchase silver more rapidly. Payment for the silver was made in Treasury notes of 1890, a legal tender currency redeemable on demand into gold or silver coin at the option of the government. The law required the purchase of 4.5 million ounces of silver per month at market prices until the market price reached one dollar for 371.25 grains of silver. Had this law remained in effect indefinitely, it would have increased the domestic credit (silver) component of the money base until only silver coin or notes convertible to silver would have been delivered to satisfy nominal contracts. Effectively, this would have been equivalent to a resumption of free coinage of silver, thereby threatening the continued circulation of gold. Within three years, the gold dollar was sufficiently threatened that the Sherman Silver Purchase Act was repealed in the midst of the Panic of 1893. However, the agitation for free coinage of silver was not extinguished until the presidential election of 1896.

II. Evaluating a Future Dollar Delivery

Here I produce a pricing formula for contracts promising future dollar deliveries in a bimetallic standard, emphasizing that a delivery option is the essential feature of a bimetallic standard. The dollar contract pricing solution and the extent of the discount will be developed in a continuous time environment. In this section, all prices will be denominated in terms of units of gold. Units of gold and silver are defined as the legal pure gold and silver content of the dollar, respectively. I will assume that an individual's portfolio may hold positions in contracts for future dollar deliveries, in silver or gold bonds, and in options to supply a unit of gold in exchange for a unit of silver in the future (silver option). Alternatively, it may contain an option to supply a unit of silver in exchange for a unit of gold (gold option).[9] These metal exchange options are standard call options. For example, since the holder of the silver option can require the delivery of a unit of silver for a unit of gold, he has a call option on silver with a striking price of unity.

---

[6] The greenbacks, introduced in 1862, were also legal tender; but by 1879 they were redeemable in coin under the Resumption Act of 1875.

[7] There is some evidence that silver was demonetized in anticipation of the decline in its relative price. See, for example, P. O'Leary (1960).

[8] Laughlin argued that the act of 1873 simply ratified a demonetization of silver that had occurred with the 1853 removal of full legal tender status for subsidiary silver coins. Since few silver dollars had been minted through 1873, he claimed that the act of 1873 merely recognized existing conditions (p. 92). Friedman and Schwartz reiterated Laughlin's argument in their discussion of the episode (p. 114).

The procedure involved in finding the time t value of a contract to deliver one dollar at time $t+\theta$ borrows directly from concepts in the option pricing literature. I find appropriate positions in pure discount gold bonds, silver bonds, or metal call options in the time t portfolio which perfectly duplicate the payoffs of the dollar contract in all future contingencies. The value of the dollar contract can be determined as the equity value of the bond and option positions through a simple arbitrage argument.

Specifically, the time t market value of an agreement to deliver one dollar at time $t+\theta$ can be derived by evaluating a position in pure discount silver bonds and silver call options which duplicates the $t+\theta$ payoff of the dollar contract. The silver bonds mature at $t+\theta$ and have a face value of one unit of silver, and the call option expires at $t+\theta$.

Using Table 1, it is easy to verify that a portfolio long one silver discount bond and short one silver call option mimics the dollar contract. One silver discount bond will pay one unit of silver at $t+\theta$, with a gold value of $P_s(t+\theta)$. The short position in the silver option will entail a zero payoff if $P_s(t+\theta) < 1$ and a negative payoff $1-P_s(t+\theta)$ if $P_s(t+\theta) > 1$. When the silver dollar is worth less than the gold dollar, the dollar contract will pay out one unit of silver with a value of $P_s(t+\theta)$. Otherwise, it will pay out a gold dollar valued at 1. The net payoff of the silver bond-silver option position is $P_s(t+\theta)$ if $P_s(t+\theta) < 1$ and 1 if $P_s(t+\theta) > 1$, which is identical to the payoff pattern of the dollar contract. Therefore, arbitrage would prevent the current gold value, $V(P_s(t),\theta)$, of the dollar contract from differing from the value of the equity in the silver bond-silver option position.

TABLE 1—PAYOFFS FROM SILVER BOND, SILVER OPTION, AND DOLLAR CONTRACT

	Contingencies at $t+\theta$	
Asset	$P_s(t+\theta)<1$	$P_s(t+\theta)>1$
1. Silver Bond	$P_s(t+\theta)$	$P_s(t+\theta)$
2. Silver Option	0	$1-P_s(t+\theta)$
3. Dollar Contract	$P_s(t+\theta)$	1

The equity value of the position depends on the assumed stochastic laws of motion for the $P_s(t)$ and for the yields on gold and silver bonds. For example, I will assume that gold and silver bonds pay continuously compounded, constant own yields π_g and π_s, respectively. Also, the relative price of silver in terms of gold is driven by the stochastic process:

$$dP_s(t)/P_s(t) = \mu\, dt + \lambda\, dZ(t), \tag{1}$$

where μ is the instantaneous expected percentage relative price change of silver; λ is the associated instantaneous standard error, assumed constant; $Z(t)$ is a Wiener process. The time t prices of gold, 1, and silver, $P_s(t)$, will move exogenously to the market for dollar contracts.

In this case the equity value of the silver bond-silver option position is

$$\text{(2a)}\quad V(P_s(t),\theta) = \exp[-\pi_s\theta]P_s(t) - D_s(P_s(t),\theta),$$

---

[9] Margrabe has priced an option to exchange one asset for another. I consider this kind of option because Stulz shows that the current value of a contract to deliver the lesser valued of assets A and B in the future equals the current price of A less the current price of an option to exchange B for A in the future. Stulz developed a pricing formula for a European call option on the less valuable of two risky assets in a continuous time framework. A European call option can be exercised only on the expiration date. For the special case in which the striking price is zero, such an option is a contract to deliver one of two assets, where the short position has the option of which asset to deliver. The option will always be exercised at the expiration date, and the short side will always deliver the less valuable of the two assets. This is exactly the contractual arrangement of an agreement to deliver future dollars in a bimetalllic standard. If the future delivery is simply the repayment of a previous loan, then the special case of a zero striking price materializes.

where $\exp(-\pi_s\theta)$ is the current value in silver of the bond, and $D_s(P_s(t),\theta)$ is the value of the silver call option expiring at $t+\theta$, given time t prices. Using a similar argument based on positions in gold discount bonds and gold options, the dollar contract's value can also be expressed as

$$\text{(2b)}\quad V(P_s(t),\theta) = \exp[-\pi_g\theta] - D_g(P_s(t),\theta),$$

where $D_g(P_s(t),\theta)$ is the value of an option to deliver a unit of silver in exchange for a unit of gold.

To derive the price $D_s(P_s(t),\theta)$, I argue that the option on silver is analogous to a call option on a stock which continuously pays a dividend equal to a constant percentage of the stock's market price. Since Merton has produced a pricing formula for such an option, I can then substitute it for D_s in (2a).[10] Since the numeraire gold will be delivered in payment by the holder in exercising the option, instantaneous gold loans are the analogs to the riskless nominal loans in the standard option pricing formulas; and the instantaneous gold yield π_g is the analog to the yield on risk-free loans.

To complete the analogy, I must consider which asset constitutes the underlying asset of the option. Since the option is a call option on one unit of silver, I must find a silver-denominated asset which delivers one silver unit at $t+\theta$. This delivery requirement is satisfied by the holding of one silver coupon bond maturing at $t+\theta$ with a par value of one unit of silver and continuous coupon rate of π_s. The silver coupon bond would deliver the one unit of silver at time $t+\theta$, so a call option on the bond exercisable at $t+\theta$ is a call on a unit of silver. Therefore, the coupon bond can represent the underlying security of the option. Also, the time t price in silver of this coupon bond is one silver dollar; so its time t price and coupon payment in gold are $P_s(t)$ and $\pi_s P_s(t)$, respectively. Since the stochastic processes driving the gold price of the silver coupon bond and its interest payments are exactly those assumed by Merton, his pricing formula is applicable:

$$\text{(3)}\quad D_s(P_s(t),\theta) = \exp[-\pi_s\theta]P_s(t)N(d_1) - \exp[-\pi_g\theta]N(d_2),$$

where $d_1 = \left[\ln(P_s(t)) + \left(\pi_g - \pi_s + \lambda^2/2\right)\theta\right]/\lambda\sqrt{\theta}$

and $d_2 = d_1 - \lambda\sqrt{\theta}$.

$N(x)$ is the value of the cumulative standard univariate normal distribution function evaluated at x, and λ^2 is the variance of the percentage change in the relative price between the gold and silver content of the dollar.

Substituting from (3) into (2) completes the calculation of $V(P_s(t),\theta)$. More generally, bimetallic coupon bonds which make discrete dollar payments $H(\theta)$ at $\theta = 1,2,3,\ldots,N$, have a current value of

$$\text{(4)}\quad VB(P_s(t)) = \sum_{\theta=1}^{N} V(P_s(t),\theta)H(\theta).$$

III. The Nominal Values of U.S. Bimetallic Securities

Here, I apply the dollar contract pricing formula derived in Section II to compute the theoretical value of bonds promising to pay bimetallic dollars issued by the U.S. government in the nineteenth century. I employ formula (4) except that I compute security values in nominal rather than in gold terms. I simply multiply both sides of (4) by $\hat{P}_g$, the nominal value of the gold dollar. For P_s, the price of the silver dollar in terms of gold dollars, I substitute $\hat{P}_s/\hat{P}_g$, where $\hat{P}_s$ is the nominal value of the silver dollar. $\hat{P}_s$, $\hat{P}_g$, π_s, and π_g are obtained from actual observations. I report the steps required to compute $\hat{P}_g$ and $\hat{P}_s$ in the Appendix. For π_g, I use the beginning of January yield on British 3% Consols from 1818 to 1882 and 2.5 percent annuities from 1883 to 1896. Since Great Britain adhered to a monometallic gold standard for most of the period and paid these loans in gold throughout the period, the Consol and annuity yields represent the yield on pure gold loans. For π_s, I used the beginning of January yield on various long-term Prussian loans through 1870. Along with the rest of the German states, Prussia maintained a pure silver standard through 1873, so these yields represent the yield on pure silver loans. However, though Prussia switched to a gold standard in 1873, it had begun a transition from silver in 1871. Therefore, for 1871–96, I used the yield on rupee loans selling in London for π_s.[11]

---

[10] For a simplified discussion of this formula, see Clifford Smith (1976, p. 26).

The assumptions in the Section II example of constant π_g and π_s and on the form of the process driving $P_s(t)$ are very strong. A glance at Table 2 indicates that π_g and π_s fluctuated from 1818 to 1896. Thus, it is desirable to construct the price of the bimetallic bond in an environment of random movement of silver and gold yields to produce a proper model of the term structures for gold and silver loans. John Cox, Jonathan Ingersoll, and Stephen Ross (1985a,b) have recently provided a method of pricing discount loans and options on bonds in a one-good optimizing model. From assumptions on the stochastic process driving the productivity of capital, they derived the equilibrium stochastic process driving the yield on instantaneous riskless loans along with the equilibrium term structure on discount bonds. Cox et al. show that arbitrage methods for pricing securities may produce prices that are not supported in equilibrium because of arbitrary assumptions about the stochastic process driving the interest rate on riskless loans. For implementation in the present problem, their model must be extended to an environment of multiple, storable goods so that silver and gold can be priced along with consumption goods. Also, the empirical implementation of their methods requires the estimation of taste parameters, the identification of which is achieved by assuming that there are no taste disturbances. On this problem, see my paper with Robert King (1984). Finally, empirical implementation of the Cox et al. model requires observations of more maturities along the term structure than are likely to be available in the present problem.

Oldrich Vasicek (1977) provides an intermediate method for addressing the stochastic nature of π_s and π_g. Using an arbitrage approach, he assumes that the instantaneous riskless interest rate follows an Ornstein-Uhlenbeck process which drifts to some long-term mean. With this assumption, he can solve explicitly for the price of discount bonds of arbitrary maturity. Empirical implementation requires time-series observations on the rate of interest on riskless loans and observations on the slope of the term structure of interest rates at the short maturities. However, there is a problem in gathering the necessary data. For example, throughout the nineteenth century, British debt consisted almost exclusively of the 3% Consols, a long-term debt instrument. A small percentage consisted of Exchequer bills. According to Charles Fenn (1883), Exchequer bills comprised 4.3 percent of the total debt in 1825, 3.5 percent in 1835, 2.1 percent in 1845, and 2.8 percent in 1855, part of which was always held by the Bank of England. Prices for new bills are available in British periodicals. However, since there was a thin market even for new bills, yield data on short bills and on the slope of the short end of the term structure would not be easily available. The problems are yet more difficult for yields on short-term silver loans; I have not yet found data on Prussian bill yields, though Sidney Homer (1977) does present data on short-term commercial loans.

In the absence of the necessary data for the more realistic pricing formulas, I have chosen to implement the Section II pricing solution. I estimated λ^2 for each year by computing a moving sample variance of actual percentage changes in the relative value of the gold dollar to the silver dollar. In the results which I report here, the sample consists of the ten previous years' percentage changes. I have also constructed estimates

---

[11] In the Data Appendix I discuss the nature of Indian rupee bonds and their rate of return.

TABLE 2—DATA EMPLOYED IN BIMETALLIC BOND FORMULA

	Dollar Price of		Yield		
Year	Gold	Silver	Gold	Silver	$\lambda^{2\,a}$
1818	1.0235	1.0000	0.0369	0.0598	0.00079
1819	0.9984	1.0000	0.0388	0.0558	0.00082
1820	1.0141	1.0000	0.0446	0.0586	0.00086
1821	1.0567	1.0000	0.0430	0.0570	0.00103
1822	1.0567	1.0000	0.0390	0.0539	0.00089
1823	1.0567	1.0000	0.0378	0.0541	0.00088
1824	1.0567	1.0000	0.0347	0.0445	0.00027
1825	1.0349	1.0000	0.0318	0.0446	0.00031
1826	1.0392	1.0000	0.0369	0.0467	0.00031
1827	1.0567	1.0000	0.0380	0.0455	0.00031
1828	1.0392	1.0000	0.0363	0.0438	0.00035
1829	1.0567	1.0000	0.0345	0.0398	0.00028
1830	1.0567	1.0000	0.0320	0.0435	0.00028
1831	1.0567	1.0000	0.0371	0.0425	0.00012
1832	1.0567	1.0000	0.0364	0.0428	0.00012
1833	1.0747	1.0000	0.0347	0.0412	0.00015
1834	1.0567	1.0000	0.0339	0.0400	0.00018
1835	1.0000	1.0095	0.0330	0.0393	0.00054
1836	1.0000	1.0116	0.0327	0.0393	0.00053
1837	1.0000	1.0180	0.0339	0.0388	0.00048
1838	1.0000	1.0086	0.0332	0.0389	0.00049
1839	1.0000	1.0256	0.0323	0.0386	0.00045
1840	1.0000	1.0256	0.0332	0.0386	0.00045
1841	1.0000	1.0235	0.0334	0.0383	0.00045
1842	1.0000	1.0129	0.0335	0.0384	0.00048
1843	1.0000	1.0057	0.0317	0.0387	0.00044
1844	1.0000	1.0044	0.0309	0.0346	0.00043
1845	1.0000	1.0086	0.0300	0.0357	0.00005
1846	1.0000	1.0044	0.0316	0.0372	0.00005
1847	1.0000	1.0192	0.0320	0.0379	0.00007
1848	1.0000	1.0002	0.0350	0.0440	0.00010
1849	1.0000	1.0129	0.0337	0.0395	0.00009
1850	1.0000	1.0107	0.0309	0.0414	0.00009
1851	1.0000	1.0447	0.0310	0.0392	0.00019
1852	1.0000	1.0298	0.0308	0.0369	0.00020
1853	1.0000	1.0404	0.0299	0.0384	0.00020
1854	1.0000	1.0447	0.0322	0.0421	0.00020
1855	1.0000	1.0447	0.0330	0.0408	0.00020
1856	1.0000	1.0425	0.0345	0.0425	0.00020
1857	1.0000	1.0510	0.0317	0.0428	0.00019
1858	1.0000	1.0425	0.0318	0.0412	0.00015
1859	1.0000	1.0171	0.0309	0.0414	0.00021
1860	1.0000	1.0510	0.0313	0.0406	0.00030
1861	1.0000	1.0383	0.0324	0.0393	0.00022
1862	1.0299	1.0651	0.0329	0.0389	0.00021
1863	1.3362	1.3931	0.0324	0.0398	0.00020
1864	1.5162	1.5808	0.0328	0.0384	0.00020
1865	2.2599	2.3562	0.0333	0.0391	0.00020
1866	1.4324	1.4965	0.0344	0.0413	0.00020
1867	1.3437	1.3867	0.0332	0.0422	0.00021
1868	1.3349	1.3663	0.0326	0.0433	0.00021
1869	1.3512	1.3887	0.0324	0.0497	0.00015
1870	1.2187	1.2499	0.0325	0.0492	0.00004
1871	1.1087	1.1371	0.0326	0.0414	0.00003
1872	1.0974	1.1279	0.0323	0.0394	0.00003
1873	1.1212	1.1369	0.0325	0.0387	0.00003
1874	1.1049	1.0864	0.0326	0.0392	0.00010
1875	1.1237	1.0954	0.0326	0.0369	0.00009
1876	1.1274	1.0679	0.0319	0.0382	0.00013
1877	1.0687	1.0372	0.0317	0.0399	0.00023
1878	1.0000	0.9112	0.0318	0.0413	0.00052
1879	1.0000	0.8412	0.0315	0.0417	0.00089
1880	1.0000	0.8889	0.0307	0.0410	0.00139
1881	1.0000	0.8645	0.0304	0.0393	0.00138
1882	1.0000	0.8815	0.0300	0.0382	0.00148
1883	1.0000	0.8518	0.0284	0.0388	0.00151
1884	1.0000	0.8624	0.0269	0.0394	0.00156
1885	1.0000	0.8455	0.0272	0.0392	0.00156
1886	1.0000	0.7882	0.0279	0.0399	0.00183
1887	1.0000	0.7819	0.0268	0.0411	0.00165
1888	1.0000	0.7533	0.0256	0.0394	0.00150
1889	1.0000	0.7204	0.0277	0.0401	0.00118
1890	1.0000	0.7522	0.0262	0.0398	0.00102
1891	1.0000	0.8222	0.0265	0.0398	0.00199
1892	1.0000	0.7416	0.0256	0.0384	0.00274
1893	1.0000	0.6473	0.0252	0.0377	0.00402
1894	1.0000	0.5382	0.0242	0.0377	0.00594
1895	1.0000	0.5255	0.0232	0.0377	0.00592
1896	1.0000	0.5191	0.0220	0.0377	0.00596

[a] Variance of percent change of relative price between gold and silver.

for λ^2 by sampling backward fifteen years and by sampling forward ten and fifteen years. The computed theoretical values of bond prices hardly vary with these changes in the method for computing estimates of λ^2. Based on estimates of λ^2, these computed values of the bimetallic bond prices are themselves estimates of the true bimetallic bond prices. All the series substituted in formula (4) are reported in Table 2.

A. *Potential Problems in Applying the Bimetallic Pricing Formula*

1. *Options Excluded from the Pricing Formula.* The pricing formula for the dollar contract was derived from the assumption that only the bimetallic option was available for the debtor. In practice, several other options were available to the government in paying off its securities; and these may cause

the observed market prices to deviate from the computed theoretical prices.

First, the government had the option of directly defaulting on its obligations. However, this was an unlikely step for a government as well established as that in the United States in the nineteenth century. More likely, the government would default indirectly either by devaluing the legal content of the dollar in terms of one or both metals or by adding a new overvalued legal tender. The U.S. government exercised both of these options in the nineteenth century in the 1834 devaluation and the 1862–78 addition of greenbacks as an overvalued legal tender. Neither of these changes affected the payment streams of existing government debt. By 1834, the federal government had paid off the national debt, so the devaluation affected only private obligations. However, to the extent that the public anticipated future possible devaluations prior to 1834, the anticipated relative price movement of the metal content of the dollar would consist of the anticipated market movement of the relative price of pure ounces of metal *and* anticipated movements in the legal ratio of silver to gold in the dollar.

The greenbacks were made legal tender for all payments *except* interest payments on the public debt, and the government continued to pay gold coin on its debt throughout the greenback era. However, since the repayment of principal with Treasury notes was not explicitly prohibited, there was extensive fear that the government would pay the principal in paper currency. This possibility ended in the March 1869 Public Credit Act, which assured future payment of principal in coin. From 1862 to 1869, however, the possibility of repayment in greenbacks forced high nominal yields on government debt and produced wide divergence between bimetallic security values and the actual prices of government debt.

In most of its bonds, the U.S. government maintained a redemption option; typically, a contract stated that the bond was redeemable at the option of the government after a given date. Rarely was there an explicit maturity. The existence of this feature made the duration of the payments stream uncertain to the lender and may therefore have affected the price of the bond. However, since most bonds were issued at high wartime coupon rates, the government almost always redeemed them immediately after reaching the contracted earliest redemption date.

In a few issues, the government gave the lender a par convertibility option into other securities. The option could be exercised on demand or, in some cases, after a certain period of time. In addition, the government gave holders of certain securities an *ex post* convertibility option, although the option was not a feature of the written contract. Because of difficulties with pricing convertible securities, I have eliminated the contractually convertible bonds from the sample.

An additional option of the government was the possibility of switching to a monometallic payment policy. Thus, it might have decided to continue delivering gold to pay off its debt even if silver became the overvalued metal. While this option would add to the cost of paying off the debt, it was the option actually exercised in the 1873 removal of the free coinage and legal tender status of silver; and it continued in force even after the restoration of legal tender silver dollars in 1878.

2. *Other Assumptions in the Pricing Formula.* In deriving the pricing formula, I assumed that the yield on pure gold and silver loans is independent of the length of the loan. However, the actual rates vary across observations. Agents, accounting for this variation, would act so that realized prices would differ from the bimetallic formula's prediction, as described above. Also, for the series on pure silver loans, I have used yields on the long-term bonds of Prussia through 1870; Prussia switched from a silver to a gold standard in the 1870's. These data may not represent the interest payments on pure silver loans if anticipations of a possible switch were built into the bond prices of the silver standard era. An additional problem with the British and Prussian yields arises from the option of these governments to redeem the bonds at par. For the British 3% Consols and the Prussian 4% loans, this would cause an over-

statement of the yield, especially in the 1840's.

Finally, the pricing formula is based on an assumption that the relative price between silver and gold is exogenous to which metal circulates in the bimetallic country. However, since some metal circulates as a medium of exchange, it is reasonable to presume that a switch in the circulating metal would feed back on the relative price. To avoid this problem requires a small country assumption for the bimetallic countries.

B. *Empirical Results: Realized Prices, Theoretical Prices, and the Value of the Bimetallic Option*

The period from 1818 to 1859 can be considered the heyday of the bimetallic dollar. Prior to 1818, the circulation of Treasury notes in the War of 1812 led to a suspension of specie payments and a divergence of the dollar from its metallic definition. By 1818, this divergence had terminated and the possibility that the government might default on bonds disappeared. Also, Great Britain and Prussia were firmly attached to the gold and silver standard, respectively. From 1834 to 1841, the government either had no debt at all or the debt consisted only of short-term notes for which no meaningful price quotations are available. By 1860, the credit of the country was in question, and from 1862 to 1869, the public feared that the government would pay principal in greenbacks. The bimetallic standard free from fears of default was restored only from 1870 to 1873, but the bill to remove free coinage and legal tender status for silver as already moving through the Congress.

In Table 3, I report actual January prices and predicted bimetallic prices in dollars for 23 different bond issues in the market from 1818 to 1896. I present the characteristics of these bonds in Table 4. Table 3 reports the theoretical value in dollars which each bond would have if its promised payments were explicitly gold dollars.

Finally, I report the dollar value of the bond's bimetallic option relative to a pure gold standard, that is, the value of an option to deliver silver in exchange for gold. I compute the option value by subtracting the bimetallic bond value from that of the gold bond. The option value represents the wealth transfer from a debtor to a creditor of a sudden switch from a bimetallic standard to a gold standard. For the period 1818–34, the option value produced theoretical discounts in dollar bonds of from 6 to 18 percent from the value of the gold bond. The 6 percent minimum discount represents the amount by which the gold dollar was undervalued at the mint at the time of the 1834 devaluation. For example, the 5% Loan of 1821 had an option value of 6.40 dollars at the time of its expiration in 1834; this amount is the value of an "in the money" option just before its expiration date. Discounts much above 6 percent occur early in the life of a bond and represent the possibility that the value of silver may decline greatly during the remaining life of the loan.

By the 1840's, there had been a 6 percent devaluation of the gold dollar and large gold discoveries which guaranteed that gold would circulate as money. In this period, the option feature of the bimetallic contract was an "out of the money" option, so the option component produced a much lower percentage discount from the value of a gold bond relative to the earlier period.[12]

C. *Some Regression Results*

From the pricing relationships among the bimetallic bond, the gold bond, and the bimetallic option emerges a regression equation:

$$B_{jt} = \phi_1 PG_{jt} - \phi_2 EG_{jt} + u_{jt}, \tag{5}$$

where B_{jt}, PG_{jt}, and EG_{jt} are the realized price, the gold bond value, and the value of

---

[12] The Loans of 1842, 1847, and 1848 expired in 1863, 1868, and 1868, respectively. The relatively high value of the option for these bonds reflects their long remaining lives and an apparent anticipated depreciation of the relative price of silver. The anticipated depreciation is reflected in a positive yield differential between silver and gold loans. The shorter-term Loans of 1843 and 1846 had very low option values in this period.

TABLE 3—BOND PRICES: ACTUAL, PREDICTED BIMETAL, OPTION VALUE, AND GOLD BOND

Year	Actual	Bimetal	Option	Gold
War Loan of 1812				
1819	99.5	101.6	9.4	111.0
1821	106.5	100.8	11.3	112.1
1822	105.5	101.5	10.4	111.9
1823	102.5	101.0	9.1	110.1
1824	102.3	101.5	6.8	108.3
War Loan of 1813				
1819	99.4	101.0	6.6	107.6
1821	107.5	100.4	8.6	109.1
1822	106.3	100.5	7.3	107.8
War Loan of 1814				
1821	107.5	101.2	13.8	115.0
1822	108.0	102.4	13.2	115.6
1823	106.0	101.9	12.3	114.3
1824	105.0	104.3	8.9	113.2
1825	106.5	102.9	6.2	109.1
1826	101.3	101.3	5.0	106.3
War Loan of 1815				
1819	100.3	102.2	12.0	114.2
1821	109.0	101.2	13.8	115.0
1822	111.5	102.4	13.2	115.6
1823	108.0	101.9	12.3	114.3
1824	108.0	104.3	8.9	113.2
1825	111.5	102.9	6.2	109.1
1826	102.3	101.3	5.0	106.3
7% Stock of 1815				
1818	109.5	105.5	17.6	123.0
1819	104.5	106.6	9.7	123.0
1820	104.0	104.4	8.4	112.8
1821	108.9	104.4	11.7	116.0
1822	107.0	104.3	10.6	114.8
1823	103.5	102.9	9.3	112.2
1824	102.3	102.5	6.8	109.3
5% Loan of 1820				
1821	102.5	93.9	18.0	111.9
1822	109.0	96.7	18.4	115.1
1823	106.0	96.8	18.5	115.3
1824	106.5	103.5	13.3	116.7
1825	110.3	103.0	12.1	115.1
1826	100.0	101.5	9.6	111.1
1827	109.3	101.8	9.5	111.3
1828	108.8	102.1	7.0	109.1
1829	101.6	102.8	7.4	110.2
1830	100.0	101.2	8.1	109.3
1831	101.0	100.7	6.3	107.0
5% Loan of 1821				
1822	109.0	95.9	21.3	117.2
1823	105.8	96.0	21.8	117.8
1824	107.0	104.5	15.6	120.2
1825	111.0	104.0	15.3	119.4
1826	100.0	102.1	12.0	114.2
1827	108.5	102.8	11.4	114.2
1828	107.5	103.5	9.1	112.5
1829	106.0	105.3	9.0	114.3
1830	105.0	102.8	11.6	114.4
1831	105.8	102.6	8.0	110.6
1832	103.8	102.0	7.7	109.7
1833	101.3	101.6	9.0	110.6
1834	100.0	101.0	6.4	107.3
4.5% Loan of 1824 (5/24/24)				
1825	105.5	100.2	11.9	112.0
1826	98.8	99.0	9.4	108.4
1827	102.0	99.7	9.3	109.0
1828	100.5	100.4	6.9	107.3
1829	98.5	101.5	7.3	108.8
1830	101.3	100.3	8.1	108.3
1831	102.3	100.2	6.3	106.5
Exchange 4.5% Stock of 1824				
1825	105.0	100.2	14.0	114.1
1826	97.8	98.8	11.0	109.7
1827	102.3	99.6	10.6	110.2
1828	103.0	100.5	8.3	108.8
1829	101.0	102.3	8.4	110.7
1830	102.5	100.5	10.3	110.8
1831	103.5	100.7	7.4	108.0
1832	100.3	100.4	7.0	107.4
1833	100.1	100.4	8.2	108.6
6% Loan of 1841				
1843	99.5	103.8	1.6	105.4
5.5% Loan of 1841				
1843	98.3	102.9	1.6	104.5
1844	100.0	101.6	0.8	102.4
Loan of 1842				
1843	101.0	128.8	12.2	141.0
1844	115.8	133.5	7.7	141.1
1845	114.5	132.6	8.5	141.1
1846	107.5	128.5	8.1	136.6
1847	101.0	127.9	6.5	134.4
1848	98.5	116.8	11.6	128.4
1849	107.0	122.7	6.0	128.7
1850	108.8	119.3	11.3	130.6
1851	112.5	123.5	5.0	128.5
1852	111.5	122.8	4.0	126.8
1853	114.3	121.1	4.5	125.6
1854	116.5	117.0	4.4	121.3
1855	116.0	116.2	2.4	118.6
1856	110.0	113.4	2.1	115.5
1857	110.0	113.0	2.2	115.2
1858	113.0	111.5	1.3	112.8
1859	102.5	108.0	2.8	110.8
Loan of 1843				
1844	103.6	111.1	4.2	115.4
1845	103.0	111.0	3.7	114.7
1846	98.5	108.5	3.5	112.0
1847	91.5	108.3	2.0	110.3
1848	91.5	102.6	4.6	107.2
1849	99.3	105.1	1.6	106.6
1850	100.1	103.6	2.5	106.2
1851	102.0	104.3	0.2	104.5
1852	100.9	102.6	0.1	102.8
1853	100.0	101.0	0.0	101.0
Loan of 1846				
1848	96.8	111.4	7.4	118.8
1849	104.5	114.8	3.2	118.0
1850	106.5	112.0	5.9	118.0
1851	107.0	114.0	1.5	115.6
1852	105.0	112.0	1.3	113.3
1853	108.5	110.3	0.9	111.2

(Continued)

TABLE 3—(Continued)

Year	Actual	Bimetal	Option	Gold
1854	106.5	107.4	0.4	107.8
1855	102.0	105.0	0.1	105.1
1856	102.0	102.5	0.0	102.5
Loan of 1847				
1848	99.3	120.4	14.7	135.0
1849	108.3	128.0	8.1	136.0
1850	111.0	124.1	15.4	139.4
1851	115.3	129.6	8.0	137.6
1852	116.3	129.9	6.3	136.2
1853	119.8	128.0	7.3	135.8
1854	121.1	122.7	8.1	130.8
1855	116.0	122.8	5.1	128.0
1856	116.3	119.6	4.9	124.5
1857	116.0	119.5	6.3	125.8
1858	112.5	118.9	4.8	123.7
1859	114.0	115.2	7.3	122.0
Loan of 1848				
1849	109.0	128.4	8.4	136.8
1850	112.3	124.5	15.8	140.2
1851	115.4	130.2	8.3	138.5
1852	117.0	130.6	6.6	137.1
1853	119.5	128.6	8.1	136.7
1854	121.3	123.2	8.5	131.6
1855	116.0	123.4	5.4	128.8
1856	116.3	120.2	5.2	125.3
1857	116.0	119.6	6.7	126.8
1858	112.0	119.6	5.2	124.7
1859	111.0	115.8	7.7	123.5
Mexican Indemnity Stock				
1850	100.8	102.4	1.2	103.6
1851	98.0	101.8	0.0	101.8
Loan of 1858				
1859	104.3	110.6	11.7	122.4
1866	96.5	155.6	2.8	158.4
1867	102.5	143.7	4.4	148.1
1868	108.0	140.6	5.3	145.9
1869	115.0	138.5	7.4	145.8
1870	110.0	125.0	4.7	129.7
1871	105.0	115.8	0.5	116.3
Loan of July and August, 1861				
1868	109.6	158.6	14.1	172.7
1869	112.3	151.0	21.8	172.7
1870	117.8	135.9	17.5	153.4
1871	110.9	130.8	6.6	137.4
1872	115.6	126.2	7.9	134.1
1783	119.1	130.6	3.8	134.6
1874	118.4	122.9	7.3	130.2
1875	118.3	123.8	6.2	130.0
1876	119.6	118.1	10.4	128.4
1877	114.1	112.0	7.3	119.3
1878	106.6	98.4	13.0	111.4
1879	106.4	87.6	19.1	106.7
1880	104.3	91.3	13.0	104.2
1881	101.5	87.3	14.1	101.5
5% Loan of 1881				
1872	112.4	121.5	3.5	125.1
1873	115.1	122.5	3.7	126.1
1874	112.9	115.9	6.8	122.7
1875	113.8	117.4	5.9	123.2
1876	116.9	112.7	9.8	122.5
1877	112.0	107.7	6.9	114.6
1878	105.3	95.4	12.3	107.7
1879	107.0	85.6	18.4	104.0
1880	103.4	89.8	12.5	102.3
1881	101.5	86.7	13.8	100.5
4% Loan of 1907				
1878	101.8	88.8	26.6	115.3
1879	99.5	81.4	34.2	115.6
1880	103.0	87.0	29.8	116.8
1881	112.5	87.1	30.1	117.2
1882	117.6	90.2	27.3	117.5
1883	119.5	86.5	33.8	120.2
1884	123.8	86.7	35.9	122.6
1885	121.8	85.3	36.0	121.3
1886	123.0	78.7	40.7	119.3
1887	127.5	76.8	43.8	120.6
1888	126.0	75.7	46.1	121.9
1889	126.0	71.8	45.8	117.6
1890	126.3	75.2	44.0	119.2
1891	121.0	84.2	33.7	117.9
1892	117.0	75.4	42.9	118.3
1893	113.5	66.2	51.6	117.8
1894	113.0	55.0	63.0	118.1
1895	113.3	53.7	64.5	118.1
1896	110.0	52.9	65.2	118.2
Loan of 1904				
1895	117.0	57.4	64.3	121.7
1896	113.3	56.2	64.2	120.4
Loan of 1925				
1896	116.3	53.8	84.5	138.4

the option to acquire a unit of silver in exchange for a unit of gold, respectively, of bond j at time t. If the option features other than the bimetallic option are of minor importance in determining the bond's market value, then $\phi_1 = \phi_2 = 1$. Carrying out an ordinary least squares regression for a first-differenced version of equation (5) for all bond prices observed for the periods 1818–34, 1842–59, and 1818–59, I report the results in Table 5.

For all the gold bond-gold option estimates, the coefficients of the exchange options differ significantly from predicted values. This divergence of the parameter estimates from the theoretical predictions is not atypical of results in the empirical options pricing literature. For example, Gay

TABLE 4—CHARACTERISTICS OF BONDS IN THE SAMPLE

Name of Bond	Coupon Rate[a]	Payment Period	Date Redeemable
7% Stock of 1815	7	Quarterly	January 1825
War Loan of 1815	6	Quarterly	January 1827
War Loan of 1812	6	Quarterly	January 1825
a. 6% Loan of 1812			
b. Exchanged 6% Stock of 1812			
War Loan of 1813	6	Quarterly	January 1823
a. 16 Million Loan of 1813			
b. 6 Million Loan of 1813			
c. Undesignated Loan of 1813			
War Loan of 1814	6	Quarterly	January 1827
a. 10 Million Loan of 1814			
b. 6 Million Loan of 1814			
c. Undesignated Loan of 1814			
5% Loan of 1820	5	Quarterly	January 1832
5% Loan of 1821	5	Quarterly	January 1835
4.5% Loan of 1824 (5/24/24)	4.5	Quarterly	January 1832
Exchange of 4.5% Stock of 1824	4.5	Quarterly	January 1834
5.5% Loan of 1841	5.5	Quarterly	January 1845
6% Loan of 1841	6	Quarterly	January 1845
Loan of 1842	6	Semiannual	January 1863
Loan of 1843	5	Semiannual	January 1853
Loan of 1846	6	Semiannual	December 1856
Mexican Indemnity Stock	5	Semiannual	January 1852
Loan of 1847	6	Semiannual	January 1868
Loan of 1848	6	Semiannual	July 1868
Loan of 1858	5	Semiannual	January 1874
Loan of July and August, 1861 (Includes Loan of 1863)	6	Semiannual	July 1881
5% Loan of 1881	5	Quarterly	April 1881
4% Loan of 1907	4	Quarterly	July 1907
Loan of 1904	5	Quarterly	January 1904
Loan of 1925	4	Quarterly	January 1925

[a] Shown in percent.

and Manaster, in regressions of futures prices on current wheat prices plus storage costs and on the delivery quality option value, find that their estimates diverge significantly from the theoretical prediction. In the bimetallic case, other option features may have been important in these bonds in the pre-Civil War period. Alternatively, a bias built into the option value estimate may affect the results.[13]

TABLE 5—RELATION OF BOND PRICE TO GOLD BOND AND GOLD EXCHANGE OPTION PRICES[a]

Period	ϕ_1	ϕ_2	R^2	Observations
1818–34	.45 (.23)	-.22 (.24)	.12	56
1842–59	.61 (.23)	.51 (.18)	.18	56
1818–59	.61 (.16)	.28 (.14)	.12	110

[a] Standard errors are shown in parentheses.

TABLE 6—RELATION OF BOND PRICE TO GOLD BOND AND GOLD EXCHANGE OPTION PRICES 5% LOAN OF 1821 AND LOAN OF 1842[a]

Period	ϕ_1	ϕ_2	R^2	Observations
1822–34	.86 (.59)	.09 (.53)	.17	12
1842–59	.72 (.48)	.92 (.42)	.33	16
1822–59	.86 (.36)	.63 (.32)	.24	28

[a] Standard errors are shown in parentheses.

Finally, since many bond observations are drawn from the same years, the disturbances in (5) will covary across bonds. To address this problem, I have isolated the observations from the 5% Loan of 1821 and the Loan of 1842 and repeated the estimation procedure reported in Table 5 for these bonds alone. I have selected these loans because each provided the longest time-series of price observations for any single bond in the periods 1818–34 and 1842–59, respectively. Table 6 contains the parameter estimates for these bonds.

---

[13] Since they are nonlinear functions of the unbiased λ^2 estimates, the bimetallic bond estimates are generally biased, as shown by John Butler and Barry Schachter (1983a, b; undated). Typically in the options pricing literature, no correctives are applied to remove the bias. Though Butler and Schachter have provided such a corrective, I have not applied it at this stage, so the results reported in this section, though consistent with standard practice in the options pricing literature, are biased.

Table 7—Homer's Reported Yields and Yields Corrected for Option Feature

Year	3% Consols	Prussian Loans	U.S. Loan of 1848	1848 Gold	1848 Silver	Value of Silver Option
1850	3.11	4.10	4.58	3.53	4.68	0.05
1851	3.09	3.99	4.47	3.88	4.78	1.04
1852	3.02	3.76	4.39	3.90	4.59	1.02
1853	3.07	3.82	4.02	3.43	4.33	0.89
1854	3.27	4.21	4.14	3.47	4.51	0.80
1855	3.31	4.10	4.18	3.77	4.54	1.22
1856	3.22	4.15	4.11	3.64	4.47	1.09
1857	3.27	4.23	4.30	3.65	4.81	0.87
1858	3.10	4.15	4.32	3.76	4.76	0.78
1859	3.15	4.29	4.72	3.81	4.94	0.24

Source: For the second, third, and fourth columns, Homer, Tables 19, 32, and 41.

For these loans, *F*-tests of the parameter restrictions support the hypothesis that $\phi_1 = \phi_2 = 1$ at standard significance levels. The difference from the previous results arises primarily because of a doubling in the standard errors of the parameter estimates, though the point estimates of the parameters are closer to the theoretical value of unity.

IV. Cross-Country Comparisons of Antebellum Yields

The results of Sections II and III are useful in addressing a line of research on comparative technological development whose conclusions hinge partly on cross-country comparisons of yields on antebellum securities. Following up on the work of H. J. Habakkuk (1962), Peter Temin (1966) examined the apparently greater use of labor-saving machinery in the United States relative to that in Great Britain during the 1850's. Choice of a capital-intensive technique would follow from a higher wage-rental ratio within the same technology or from a different technology in which capital was relatively more productive in the United States. Based on observations in Homer of higher yields on securities in the United States, Temin argued that the wage-rental ratio was actually higher in Great Britain. Therefore, the observed relative factor intensities conflicted with an assumption that both countries had access to the same technology.

Edward Ames and Nathan Rosenberg (1968) and Alexander Field (1983) developed this cross-country comparison by considering more refined production technologies while accepting the claim that a higher observed yield in the United States was associated with a higher cost of capital. Paul Uselding (1972) and V. Kerry Smith (1978) turned in the direction of estimating sophisticated antebellum production functions for the United States alone, but they both treated observed U.S. yields as pure interest rates.

However, since dollar securities contained an option feature not contained in sterling securities, otherwise identical debt contracts in the different nominal units should have commanded different prices. For the simple case of a pure discount loan, the bimetallic pricing equation (2b) indicates that the price of the dollar loan equals the price of the sterling loan less the value of the option to pay silver instead of gold. Since the live option always has a positive value, dollar loans must command lower market price in gold than sterling loans, even if yields on pure gold loans in the two countries are identical. By the same logic, a dollar loan should command a lower price than a pure silver loan.

These implications are supported by evidence reported in Table 7. The second, third, and fourth columns are Homer's reported yields from the 1850's on British 3% Consols, Prussian loans, and the U.S. Loan of 1848, respectively. Historians have interpreted these data as signals of a higher capital cost in the United States than in Great Britain. In computing the yields, Homer assumes that the foreign loans were perpetuities, but he calculates yields to maturity for the U.S. loans. The reported yields are annual averages.

A gold bond with characteristics otherwise identical to the Loan of 1848 should have a value equal to that of the bimetallic bond plus the value of the gold option. For each year, I have converted Homer's U.S. bond yields into the implied market price and added the computed values of the gold options for the Loan of 1848 from Table 3.[14] This produces an estimate of the market price of a Loan of 1848 payable in gold, assuming that U.S. and British gold yields and U.S. and Prussian silver yields are identical. Next, I converted these price estimates into yields to maturity, producing a series labelled 1848 Gold on U.S. gold yields reported in the fifth column of Table 7.

The U.S. yields corrected for the bimetallic option are much closer to British yields than are Homer's yields. In one case, the U.S. yield falls by more than 100 basis points. For the decade of the 1850's, average British yields were 3.16 percent. The uncorrected and corrected average U.S. yields were 4.32 and 3.68 percent, respectively. The uncorrected yield differential is 116 basis points, and the corrected yield differential is 52 basis points. The bimetallic option accounts for more than one-half the yield differential.

If the Loan of 1848 had promised to pay only silver, its value could be estimated by adding the value of the silver option (in gold) to the market price implied in Homer's yields. Dividing this sum by the market price of the silver dollar, $\hat{P}_s(t)$, produces the silver price of the U.S. silver bond. Table 7 contains computations of the gold value of the silver options for the Loan of 1848; and the column "1848 Silver" contains the corrected yields of the Loan of 1848 payable in silver.[15] I constructed these yields in a manner analogous to that of the corrected gold yields, converting the silver price of the silver Loan of 1848 into a yield to maturity. Because of the low value of the silver option in the 1850's, the corrected silver yields, though somewhat higher, are much closer to Homer's reported yields than are the corrected gold yields. The increase in yield results mainly from the conversion from gold to silver units, since $\hat{P}_s(t)$ exceeds unity in this period. For the 1850's, Prussian bond yields averaged 4.08 percent. The corrected U.S. silver bond yield averaged 4.64 percent, an increase of 32 basis points from the uncorrected average yield. The differential between U.S. and Prussian silver yields averaged 56 basis points during the decade, compared to the 52 basis-point differential between U.S. and British gold yields.

V. Option Values and the Silver Agitation

By 1873, even after a rapid depreciation of silver, the option feature of the bimetallic bond was still "out of the money." For the Loan of 1858, redeemed before the 1873 demonetization of silver, little value was attached to the silver delivery option. Even the longer-term loans, such as the Loan of July and August, 1861 and the 5% Loan of 1881, have bimetallic option values in the early 1870's comparable to those prevailing in the 1840's when there was little chance of exercise.

Using the bimetallic option values associated with these loans, we can estimate the magnitude of the transfer from debtors to creditors generated by the termination of free coinage of silver in 1873. For the Loan of July and August 1861, I calculated that the value of the option to deliver silver was 3.80 dollars. Using the theoretical bimetallic bond value of 130.60 dollars, the transfer from the debtor was 2.9 percent of the value of the bond. Using the actual bond value of 119.10 dollars, the transfer from the debtor was 3.2 percent of the value of the bond. This is of the order of magnitude of one year's interest payment. Repeating the computation for the 5% Loan of 1881 yields similar results.

---

[14] The option values were calculated based on gold and silver yields observed in January. Since Homer's yields were annual averages, some discrepancy may arise.

[15] I computed the silver option value as follows. First, I discounted the remaining payments stream of the Loan of 1848 with the Prussian yield from Table 2 and multiplied by $P_s(t)$. This produced the gold value of the Loan of 1848 as if it were a silver bond, assuming that the U.S. silver yield equalled that of Prussia. Subtracting the theoretical bimetallic bond values in gold reported in Table 3 from this silver bond value produces the gold value of the silver option implicit in the bond.

This computation of the extent of the transfer may be misleading because of the wide divergence of the theoretical bimetallic bond price from the actual market price in 1873. This divergence signals that the simple pricing formula does not capture some important features of the bond. For example, doubt about the resolve of the government to avoid paying off its debt in greenbacks would cause the theoretical dollar price of the bimetallic bond to exceed the observed price.

From the spreads between the theoretical prices of the gold bonds and the actual prices of the dollar bonds, an alternative method of estimation of the value of the option in 1873 is available. For the Loan of July and August 1861, the spread declined from 15.50 dollars in 1873 to 11.80 dollars in 1874. Attributing this decline to the removal of the option yields an option value of 3.70 dollars in 1873. Thus, the transfer was 3.1 percent of the actual value of the bond. For the 5% Loan of 1881, the transfer was 1 percent.

Only by the late 1870's, with the continued decline in silver, did the option value associated with a restored bimetallic standard rise to unprecedented levels. By this time, the dollar was de facto a gold dollar, as evidenced by the near equality between the realized bond values and the gold bond values. This equality emerges for the first time in the entire sample in the late 1870's. The bimetallic option values are indicative of the magnitude of the transfer from creditor to debtor which would have occurred with the reintroduction of free coinage of silver. By the 1890's, the potential transfer reached levels exceeding 50 percent of the realized value of the bonds. Indeed, simply comparing the market price of silver to the legal silver content of the dollar would indicate the potential transfer to debtors of a return to free coinage of silver, since the option was very much in the money by 1890. The rapidly increasing potential transfer parallels the increasing intensity of the silver agitation, which reached its peak in the mid-1890's.

VI. Conclusion

Having provided methods for pricing a wide variety of previously inaccessible assets, finance theory has led more or less naturally to the description of numerous contracts or institutions as essentially options. Though overlooked by modern analyses of bimetallism, the option of which metal to deliver encompasses the fundamental aspect of the bimetallic standard; and proponents of bimetallism in the nineteenth century recognized the central importance of the option to the system. The pricing of bimetallic securities provides a natural application of option pricing methods. Pricing the option feature of such contracts opens several issues to examination. First, one can easily show that antebellum international yield comparisons, used often in research on the nature of nineteenth-century production functions, are misleading because of the different nature of otherwise similar securities in bimetallic and monometallic standards. Second, the value of the option provides a measure of the magnitude of transfer from debtor to creditor of a switch from a bimetallic to a gold or silver standard.

DATA APPENDIX

I have used several price and yield series in computing the dollar contract prices in the text. In particular, I employ series on both the market nominal prices of and the relative price between the quantities of gold and silver legally contained in the dollar. In addition, I have used yield series on bonds requiring the delivery of only gold or only silver. Here I describe the construction of these series.

The Ratio of Fine Ounces of Gold to Fine Ounces of Silver

From 1800 to 1817, I used annual average ratios of nominal gold to nominal silver prices, as reported in Laughlin (p. 290). These were market ratios gathered by Soetebeer from trading in Hamburg.

From 1818 to 1896, I used the nominal prices of standard gold and standard silver in London in early January, as reported in *The London Times* through 1844 and in *The Economist*, 1845–96. The silver prices for 1824 and 1825 were taken from Thomas Tooke (1838, p. 385). I have used London rather than New York prices because of the greater standardization of the metals which were traded in London. Throughout the century, London was the central market for silver. Weekly New York prices for American gold coin, Spanish gold doubloons, and Spanish silver dollars are available from 1817 in the *New York Commercial*. However, as discussed in government papers written by Treasury Secretary Ingham in 1830 (see U.S. Senate, Executive Document No. 58, 1879, pp. 558–84) recommending the use of London prices to gauge the market ratio between gold and silver, the weights of the coins traded in New York varied sufficiently that one can place no faith in their consistency. By the time of the Civil War, the New York gold market had become an active and accurate indicator of the market price of standard gold. Also, by the 1870's, accurate silver prices are available from New York trading. However, to maintain consistency, I use the London prices throughout. Since most government bond prices were taken from mid-January or later, I chose early January gold and silver prices to allow for the transmission of information about metal price movements to New York.

To begin the construction of the series, let p'_g and p'_s represent the nominal price in pence of an ounce of fine gold and silver, respectively. From 1822 onward, Great Britain maintained a gold standard in which one ounce of gold, 11/12 fine, had a nominal value of 934.5 pence. Since a standard ounce of silver was 37/40 fine, Laughlin (p. 295) provides a formula for producing the market ratio of the number of ounces of fine silver to an ounce of fine gold: $p'_g/p'_s = 943/x$ where x is the current price of standard silver in pence.

For 1818–21, the nominal price of gold varied, and there was no direct market quote for standard gold. However, *The London Times* provided prices for "Foreign Bar Gold." I presumed that foreign bar gold was standard gold. The price of standard gold multiplied by 12/11 is the price of fine gold, that is, $p'_g =$ [(Foreign Bar Price in Pence)]$\times 12/11$. Then the ratio of fine silver to fine gold for 1818–21 is $p'_g/p'_s = p'_g/[x\ 40/37]$. I will refer to an element of this series as q_t.

The Nominal Values of the Gold and Silver Content of the Dollar

From the q_t series, I computed the nominal prices $\hat{P}_g$ and $\hat{P}_s$ of the legal amount of gold and silver in the dollar, respectively. For the entire sample, the legal silver content of the dollar was .7733 fine ounces. Through 1834, the legal gold content of the dollar was 24.75 grains or .0515625 fine ounces. From 1835 to 1837, the legal gold content was 23.2 grains or .048333 fine ounces; and from 1838 onward, the legal gold content of the dollar was 23.22 grains or .048375 fine ounces.

Until 1834, silver was overvalued at the mint, so silver coins circulated in payment of dollar obligations. The nominal price of the silver dollar, $\hat{P}_s$, was therefore one dollar; and the nominal value of an ounce of fine silver was 1.293 dollars. The nominal value of the gold dollar can be determined from the market relative price between gold and silver, the nominal price of silver, and the gold content of the dollar. Through 1834, $\hat{P}_g = 1.293 \times q_t \times .0515625 = 0.06667 q_t$.

From 1835 to 1861, the gold dollar circulated as money, so the nominal price of the gold dollar was one dollar. From 1835 to 1837, the nominal price of a fine ounce of gold was therefore 20.689669 dollars; while from 1838 to 1861, the nominal value of a fine ounce was 20.671835 dollars. The nominal value of the silver content of the dollar was $\hat{P}_s = 20.6897(.7733)/q_t = 16./q_t$ from 1835 to 1837; and from 1838 to 1861, $\hat{P}_s = 20.67184(.7733)/q_t = 15.9855/q_t$.

From 1862 to 1878, greenbacks were delivered in payment of dollar contracts (except government debt payment and explicit gold loans). However, an active gold exchange began in January 1862, so the nominal market price of the gold dollar $\hat{P}_g$ is directly available on a daily basis for this period. I have taken price data from Wesley Mitchell (1903) and *Commercial and Financial Chronicle* to coincide with the dates for which bond prices are reported. The nominal price of the silver dollar is $\hat{P}_s = .7733\hat{P}_g/(.048375 q_t) = 15.9855 P_g/q_t$.

From 1879 to 1896, gold again circulated so the nominal price of the gold dollar was one dollar and that of the silver dollar was $\hat{P}_s = 15.9855/q_t$. For the period 1818–96, I report the nominal prices of the gold and silver content of the dollar in Table 2.

Bond Prices and Characteristics

For the prices of U.S. government bonds from 1818 to 1859, I have used January price quotes from *New York Commercial* with a few exceptions. For the 1819 prices of the War Loans of 1812, 1813, and 1815, I used *Grotjan's Philadelphia Public Sale Reports'* listings for the Baltimore Stock Market for January 18, 1819. I used the *Saturday Evening Post* (Philadelphia) for Philadelphia dealers' quotes for Mexican Indemnity Stock (January 19, 1851) and for 1855–57 quotes for the Loans of 1842, 1846, 1847, and 1848. From 1866 to 1896, I have used January price quotes from *Commercial and Financial Chronicle*. When quotes were listed as a bid-asked spread, I let the average represent the price. Since some outstanding securities were not listed every day, I could not always employ price quotations from the same day. Most often, I used prices in the middle of January, but for some issues in some years, I had to employ price information from the end of January or the beginning of February. In Table 3, I list the prices of some bonds used in the sample years. In Table 4, I list the coupon and redemption characteristics of these bonds; this information was taken from R. A. Bayley (1881) and William DeKnight (1900).

I did not include every bond issued as part of the sample For some issues the repayment period was completely at the discretion of the government, so they were hard to price. Therefore, I did not include such loans as the 5.5% Stock of 1795, the 4.5% Stock of 1795, the Navy 6% Stock, The Converted 6% Stock of 1807, the Temporary Loan of 1812, the Mississippi Stock, the 5% Loan of 1816, and the Bounty-Land Scrip. For Treasury notes, the issue and redemption dates were uncertain, so I could not interpret the nature of the payment stream represented by a given price. Thus, I had to eliminate short-term debt issued in the War of 1812, the 1837–40 depression, the Mexican War, and the Civil War from the sample. Also eliminated are the 5–20's of the Civil War, the Consols of 1865, and the Consols of 1868 because of the uncertain redemption time associated with these issues. Finally, the 7–30's of 1864 and 1865 were eliminated because they contained a contractual convertibility feature and because they were payable in greenbacks. Included in the sample are 23 of the approximately 125 government securities issued between 1790 and 1895.

To derive the yield on a pure gold loan, I used the early January London ex-coupon prices of 3% Consols from 1818 to 1882. I found these prices in issues of *The London Times* from 1818 to 1845 and in *The Economist* from 1846 to 1882. After 1882, I used the London prices of 2.5 percent annuities as reported in *The Economist* to compute the gold loan yield because the consol price was affected by an increasing probability of par redemption as the market yield declined. See Homer (pp. 192–200) or Fenn (pp. 24–30) on this issue.

To derive the yield on pure silver loans, I used the prices of Prussian bonds from 1818 to 1870, a period when Prussia was on a pure silver standard. These yields are available in Julius Kahn (1884, pp. 209–12). From 1818 to 1843, I used the end of December prices for 4 percent government bonds to derive silver yields applicable to January of the next year. From 1844 through 1868, I used end of December prices for 3.5 percent government bonds. For the 1870 yield, I used the March 30, 1870 price of 4 percent bonds because of the termination of the 3.5 percent price series.

Since India was also on a silver standard until it shifted to a gold standard in 1894, I used the yield on rupee bonds to represent the silver loan yield for 1871–93. For 1894–96 yields, I simply used the rupee yield of 1893. India shifted from a bimetallic to a silver standard in 1835. However, since India was governed by the East India Company, rupee bonds were not traded in London. The only official external borrowing was represented by the sterling denominated debt of the East India Company. With the rebellion of 1857, the East India Company was replaced by the Imperial Government as the Government of India, and immediately this government began to finance itself by selling securities in London. Since the Government of India was a creature of the British government, its debt was considered almost as secure from default as British government debt.

Indian debt was mainly denominated in sterling, but a large amount of rupee paper traded in London. A category of rupee debt called "Enfaced Paper" was the most regularly quoted in the London Markets. The rupee coupon and principal of Enfaced Paper was payable in London as a rupee draft on Calcutta. The most often quoted of the Enfaced Paper was the 4% Enfaced Paper. Since after 1879 this instrument was payable on three months notice, it was comparable in its option features to the British 3% Consols. *The Economist* contains prices of 4% Enfaced Paper, which I used to compute silver yields. The gold and silver yields for 1818–96 are reported in Table 2.

REFERENCES

Ames, Edward and Rosenberg, Nathan, "The Enfield Arsenal in Theory and History," *Economic Journal*, December 1968, *78*, 827–42.

Barro, Robert J., "Money and the Price Level Under the Gold Standard," *Economic Journal*, March 1979, *89*, 13–33.

Bayley, R. A., *The National Loans of the United States from July 4, 1776 to June 30, 1880*, Washington: USGPO, 1881.

Butler, John S. and Schachter, Barry, (1983a) "Biased Estimation of the Black Scholes Formula," Working Paper 84–W22, Vanderbilt University, May 1983.

______ **and** ______, (1983b) "Unbiased and Mean Square Error Estimation of the Standard Deviation in the Estimation of Option Prices: A Note," Working Paper, Vanderbilt University, November, 1983.

______ **and** ______, "Unbiased Estimation of the Black-Scholes Formula," Working Paper 83–W31, Vanderbilt University, undated.

Cox, John C., Ingersoll, Jonathan E. and Ross, Stephen A., (1985a) "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, March 1985, *53*, 363–84.

______, ______, **and** ______, (1985b) "A Theory of the Term Structure of Interest Rates," *Econometrica*, March 1985, *53*, 385–408.

DeKnight, William F., *History of the Currency of the Country and of the Loans of the United States*, Washington: USGPO, 1900.

Dunbar, Charles F., *Laws of the United States Relating to Currency, Finance and Banking from 1789 to 1891*, Boston: Ginn and Company, 1883.

Fenn, Charles, *Fenn's Compendium of the English and Foreign Funds*, London: Effingham Wilson, Royal Exchange, 1883.

Field, Alexander, "Land Abundance, Interest/Profit Rates, and Nineteenth Century American and British Technology," *Journal of Economic History*, June 1983, *43*, 405–31.

Fisher, Irving, *The Purchasing Power of Money*, New York: McMillan, 1913.

Friedman, Milton, "Financial Futures Markets and Tabular Standards," *Journal of Political Economy*, February 1984, *92*, 165–67.

______ **and Schwartz, Anna J.,** *A Monetary History of the United States*, NBER Business Cycles, No. 12, Princeton: Princeton University Press, 1963.

Garber, Peter M. and King, Robert G., "Deep Structural Excavation Identification Problems in Euler Equation Methods," Working Paper, University of Rochester, July 1984.

Gay, Gerald and Manaster, Stephen, "The Quality Option Implicit in Futures Contracts," *Journal of Financial Economics*, September 1984, *13*, 353–70.

Habakkuk, H. J., *American and British Technology in the 19th Century*, Cambridge: Cambridge University Press, 1962.

Hamilton, Alexander, "Treasury Report on the Establishment of a Mint" (January 28, 1791), in H. Krooss, ed., *Documentary History of Banking and Currency in the United States*, Vol. 1, New York: Chelsea House, 1969, 162–192.

Homer, Sidney, *A History of Interest Rates*, New Brunswick: Rutgers University Press, 1963; 1977.

Jevons, William S., *Money and the Mechanism of Exchange*, New York: D. Appleton, 1899.

Kahn, Julius, *Geschichte Des Zinsfusses in Deutschland seit 1815*, Stuttgart: J. G. Cotta'schen Buchhandlung, 1884.

Laughlin, J. Laurence, *The History of Bimetallism in the United States*, New York: D. Appleton, 1885.

Margrabe, William, "The Value of an Option to Exchange One Asset for Another," *Journal of Finance*, March 1978, *23*, 177–86.

Marshall, Alfred, "Remedies for Fluctuations of General Prices" (1887), in A. C. Pigou, ed., *Memorials of Alfred Marshall*, London: Macmillan, 1925.

Merton, Robert, "Theory of Rational Option Pricing," *Bell Journal of Economics*, Spring 1973, *4*, 141–83.

Mitchell, Wesley C., *A History of the Greenbacks*, Chicago: University of Chicago Press, 1903.

O'Leary, P., "The Scene of the Crime of 1873 Revisited: A Note," *Journal of Political Economy*, August 1960, *68*, 388–92.

Rolnick, Arthur and Weber, Warren, "Gresham's Law or Gresham's Fallacy," *Journal of Political Economy*, February 1986, *94*, 185–99.

Smith, Clifford, "Option Pricing: A Review," *Journal of Financial Economics*, January/March 1976, *3*, 3–51.

Smith, V. Kerry, "The Ames-Rosenberg Hypothesis and the Rule of Natural Resources in Production Technology," *Explorations in Economic History*, July 1978, *15*, 257–68.

Stulz, René, "Options on the Minimum or the Maximum of Two Risky Assets," *Journal of Financial Economics*, July 1982, 10, 161–85.

Temin, Peter, "Labor Scarcity and the Problem of American Industrial Efficiency in the 1850's," *Journal of Economic History*, September 1966, *26*, 277–98.

Tooke, Thomas, *A History of Prices and of the State of the Circulation from 1793 to 1837*, London: Longman, Orme, Green and Longmans, 1838.

Uselding, Paul, "Technical Progress at the Springfield Armory, 1820–1850," *Explorations in Economic History*, Spring 1972, *9*, 291–316.

Vasicek, Oldrich, "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, November 1977, *5*, 177–88.

U.S. Congress, *Report of the U.S. Monetary Commission*, Washington: USGPO, 1877.

U.S. Senate, *Report of the Proceedings of the International Monetary Conference, 1878*, Executive Document No. 58, Washington: USGPO, 1879.

______, *Coinage Laws of the United States, 1792 to 1894*, Finance Committee, Washington: USGPO, 1894.

Is Increased Price Flexibility Stabilizing?

By J. BRADFORD DE LONG AND LAWRENCE H. SUMMERS*

This paper uses John Taylor's model of overlapping contracts to show that increased wage and price flexibility can easily be destabilizing because of the Mundell effect. While lower prices increase output, the expectation of falling prices decreases output. Simulations based on realistic parameter values suggest that increases in price flexibility might well increase the cyclical variability of output in the United States.

Most economists would answer the question posed in our title affirmatively. It is certainly true that if prices were completely inflexible, markets could never equilibrate to changing conditions. And given a disturbance in nominal aggregate demand, "more flexible" prices imply that more of the disturbance is absorbed as a change in the price level, leaving less to appear as a cyclical movement in production. Considerations along these lines have led to a widespread belief that an excessively high degree of wage and price inflexibility is responsible for cyclical fluctuations in employment and output.

This paradigmatic belief—that the magnitude of the business cycle depends largely on the degree of institutional wage rigidity—is institutionalized by the pedagogical style of many macroeconomics courses. At some point a "classical" aggregate supply curve generated by complete price flexibility and atomistic markets, is counterposed to a "Keynesian" aggregate demand-aggregate supply model, where the short-run aggregate supply curve is infinitely price elastic because nominal wages are assumed fixed. The teacher then demonstrates that in the classical model a monetary (or any other nominal) shock has no real effects on output because the aggregate price level adjusts to clear all markets, and, in the Keynesian model, a monetary shock has real effects because quantities adjust. The message carried away is that the crucial market failure generating business cycles is the existence of sticky nominal wages: if only the labor market were fully competitive and flexible, then business cycles would be much smaller. This view suggests the desirability of any possible policy measures which increase wage and price flexibility.

This message is too simple. We believe, as we argued in our forthcoming paper, that the sign of the macroeconomic consequences of increased aggregate price flexibility is not a settled issue. We think that the question "is increased price flexibility stabilizing?" is—or should be—kept open, and there is a real chance that aggregate price flexibility would be destabilizing at the margin given the present structure of the American economy.

The view that price level flexibility can be destabilizing has been expressed many times. Irving Fisher (1923, 1925) saw the business cycle as "largely a 'dance of the dollar' ": expected deflation led to high anticipated real interest rates, little investment, and low output. If the money supply were manipulated to stabilize the price level—said Fisher—much of the business cycle would disappear, for

> what concerns business is not whether prices are high or low but whether they are rising or falling. Thus rising prices stimulate business because the prices a

---

*Harvard University, Cambridge, MA 02138, and NBER. We have benefited from useful discussions with Olivier Blanchard, Guillermo Calvo, and Greg Mankiw, and financial support from the National Science Foundation. The research reported here is part of the NBER's research programs in Economic Fluctuations and Financial Markets and Monetary Economics. Any opinions expressed are our own and not those of the National Bureau of Economic Research.

> producer can get outrun his expenses for interest rent, salaries, and wages, while falling prices depress trade ...[W]e find that this one element, *rapidity of price movement*, during the period 1914–1922 seems to account, almost completely, for the ups and downs of business. [1923, p. 1027].

Keynes argued in his chapter 19 that it was preferable in a gathering recession to increase real balances by printing money rather than by helping along the decline in prices, because

> it would be much better that wages should be rigidly fixed and deemed incapable of material changes, than that depression should be accompanied by a gradual downward tendency of money-wages, a further moderate wage reduction being expected to signalize each increase of, say, one percent in the amount of unemployment. For example, the effect of an expectation that wages are going to sag by, say, two percent in the coming year will be roughly equivalent to the effect of a rise of two percent in the amount of interest payable for the same period. [p. 265]

And, according to Lars Jonung (1981), the Swedish Central Bank's attempt to adopt a price-stability rule for the conduct of monetary policy greatly insulated Sweden from the Great Depression.

More recently, James Tobin (1975) has stressed the potential role of destabilizing deflation in creating a situation in which there is "protracted unemployment which the natural adjustments of a market economy remedy very slowly if at all" (p. 200). He argues that "even with stable monetary and fiscal policy, combined with price and wage flexibility, the adjustment mechanisms of the economy may be too weak to eliminate persistent unemployment" (p. 200), because the destabilizing price change effect on aggregate demand may well swamp the stabilizing price level effect.

In our forthcoming paper, we noted that the increasing degree of nominal rigidity in the U.S. economy over the past century had coincided with a dramatic decrease in the cyclical volatility of economic activity. This observation tends to bely explanations of cyclical fluctuations based on nominal rigidities. We presented models with backward-looking expectations, in which marginal increases in aggregate price flexibility were destabilizing, in the sense that they increased the steady-state variance of output in response to independent stochastic shocks. We conjectured that the same property—destabilizing price flexibility—would hold true in at least some models with fully rational expectations.[1]

This paper presents such a model. We take John Taylor's 1979 and 1980 macroeconomic model with staggered labor contracts and modify it so that (a) it produces persistent fluctuations of output in response to macroeconomic demand shocks, and (b) it allows anticipated changes in the price level to affect the level of real aggregate demand through effects on real interest rates or the distribution of wealth. Solving the model numerically, we find that for a wide range of parameter values, an increase in the responsiveness of wages to excess demand or supply in the labor market leads to an increase in cyclical variability as measured by the steady-state variance of output.

The paper is organized as follows: Section I describes our modifications of the Taylor model and shows that increasing the flexibility of prices may well be destabilizing in our modified model. Section II verifies the robustness of this result by examining modifications of the time structure of contracts. We are able to verify that the volatility of output ultimately declines as the wage-price process approaches perfect Walras-like flexibility. But this limiting result is very misleading as a guide to marginal changes in price flexibility that start from current levels. Section III considers supply shocks and shows that our conclusion that price flexibility can be destabilizing holds for situations in which there are a mixture of supply and demand shocks. Section IV generalizes the analysis by allowing aggregate demand to depend on both Tobin's q and the rate of deflation as well as the real interest rate. The former generalization tends to increase the stabilizing character of price flexibility while the latter reduces it. Section V concludes the paper by discussing some empirical observations supporting our analysis and discussing directions for future research.

---

[1] Bennett McCallum (1983) presents a model in which price flexibility is never destabilizing. Our conjecture has also been challenged by Robert Driskill and Steven Sheffrin (1986). For reasons discussed below, we believe that their model does not provide a satisfactory framework for examining the issue of destabilizing price flexibility.

I. Wage Rigidity and Output Volatility

Taylor (1979 for the intuitive, simplified version; 1980 for the full model and extensions) set forth the following simple macroeconomic model to illustrate how the asynchronization of nominal price-setting decisions can have important aggregate effects:

(1) $y_t = m_t - p_t + v_t$ aggregate demand

(2) $m_t = hp_t$ money supply rule

(3) $p_t = .5(w_t + w_{t-1})$ definition of price level

(4) $w_t = .5w_{t-1} + .5E_{t-1}w_{t+1} + g(.5E_{t-1}y_t + .5E_{t-1}y_{t+1}) + e_t$ wage-setting equation

where e_t and v_t are stochastic shocks.

Equations (1) through (3) are completely standard. Equation (1) says that with variables in logarithms, output y_t is equal to money balances $m_t - p_t$ plus a white-noise velocity shock v_t. Equation (2) asserts that the monetary authority partially accommodates increases in the price level by increasing the nominal money supply in the proportion h. Equation (3) defines the aggregate price level. Taylor supposes that workers negotiate two-period fixed nominal wage contracts. The aggregate price level in period t is simply the average of the two different contracts in force at that moment—the contract which begins in period $t-1$ and the contract which begins in period t. Each contract covers half of the labor force, and w_t is defined as the (log of the) wage paid in periods t and $t+1$ to those workers whose contract is negotiated at the end of $t-1$.

Equation (4) contains the heart of the model. The contract wage w_t is set as the average of the contract wages w_{t-1} and the expected contract wage w_{t+1}—the average of the wages of those contracts which overlap the periods covered by the contract which pays w_t—adjusted for excess demand or supply in the labor market through the terms containing $E_{t-1}y_t$ and $E_{t-1}y_{t+1}$. The "E_{t-1}" appears because the contract is negotiated before the realization of period t's shocks. The parameter g represents the degree of wage flexibility: the higher is g, the greater is the response of wages to anticipated demand conditions.

Recent research, George Akerlof and Janet Yellen (1985), N. Gregory Mankiw (1985), and Olivier Blanchard (1985), suggests that contracts of the type envisioned by Taylor can be justified on microeconomic grounds. The private loss from such arrangements is second-order, relative to the social consequence which is first-order. In the presence of small transactions costs, firms and workers may therefore enter into such contracts. Alternatively, following the arguments advanced in Stanley Fischer (1984), the analytical use of a contracting model of this type may be justified on empirical grounds: contracts similar to those envisioned by Taylor do predominate in American labor markets, even if economists cannot fully rationalize their existence.

Taylor's model provides a clean and forceful implementation of the fundamental insight that the asynchronization of price changing decisions can have significant consequences for the business cycle. But for the purpose of examining the effects of varying degrees of price flexibility, Taylor's particular implementation of this idea is deficient in two major respects: first, the specific models advanced by Taylor (1979, 1980) produce the wrong kind of output persistence; second, the specific model possesses no channel through which price flexibility could be stabilizing.

As Taylor demonstrates, equations (1) through (4) can be solved to produce a moving average representation for output as a function of the shocks e_t and v_t:

$$(5)\quad y_t = .5\beta e_t - .5\beta \times \left\{ \sum_{1}^{\infty} \left((D)^{i-1} + (D)^{i} \right) e_{t-i} \right\} + v_t,$$

where D is some complex function of structural parameters and $\beta = 1 - h$. An aggregate demand shock—the velocity shock v_t—generates no persistent fluctuations. The only source of persistent fluctuations is the e_t shock in equation (4).

The business cycle described by Taylor in equation (5) is, therefore, a cost-push business cycle. Cyclical output fluctuations are driven by the e_t supply-side shock to the nominal wage level, not by the v_t shock to demand. As Taylor's model stands, it has relatively little to tell us about the persistent output fluctuations in response to demand-side shocks that Keynesian economists, at least, think lie at the heart of the business cycle.[2] If a version of Taylor's (1979) model is to generate persistent output fluctuations as a result of demand shocks, the demand shocks must be serially correlated.

For our purposes, there must also be an additional modification: as Taylor's model stands, there is no channel through which price flexibility could possibly be destabilizing. Given v_t and the monetary policy reaction function, output moves inversely with the price level. In order to model how deflation might exert downward pressure on output, we must complicate the determinants of aggregate demand. We do this by relaxing Taylor's assumption that the demand for money is interest inelastic.

The modified model exhibits a kind of Mundell effect. The interest rate that determines demand for goods is a real interest rate; the interest rate that clears the money market is a nominal interest rate. Inflation or deflation drive a wedge between the two and shift the short-run solution of the *IS-LM* system, which determines output. While a lower price level is expansionary, the expectation of falling prices is contractionary, creating the possibility of instability. This seems the easiest but not the only way to model the host of possible channels—redistributions, bankruptcies, liquidity failures, real interest rate changes, and so forth—through which deflation might depress output.

We assume output as determined by the *IS* curve depends on the short-run real interest rate and a serially correlated demand shift term s_t:

$$(6)\quad y_t = -A\{i_t - [E_t p_{t+1} - p_t]\} + s_t$$

$$(7)\quad s_t = \mu s_{t-1} + z_t,$$

where z_t is independent and identically distributed. The money-demand function is standard, with velocity depending on the nominal interest rate:

$$(8)\quad m_t = p_t + y_t - v_t$$

$$(9)\quad v_t = G i_t.$$

Paralleling Driskill and Sheffrin, the money authority follows an interest rate rule which is designed to remove m_t as an independent variable from (8) and (9):

---

[2] What economic interpretation can be given to a positive e_t shock? It is an exogenous increase in the wage rate attached to the particular contract that begins in period t. It may be thought of as representing an exogenous outburst of union militancy. These episodes of successful bargaining lead to inflation, to a higher price level. Because the monetary authority does not fully accomodate this cost-push shock, this price level increase is gradually rolled back by a prolonged period of subnormal output. The business cycle described by Taylor is, therefore, a cost-push business cycle.

Driskill and Sheffrin show that in such a model where the business cycle is driven by supply shocks, then price flexibility must be stabilizing at the margin. This creates an interesting tradeoff: we show that under some circumstances, price flexibility is destabilizing at the margin in response to demand shocks. The degree of price flexibility that minimizes the variance of output will thus depend on the relative magnitudes and the correlations of demand and supply shocks. We discuss this further below.

$$m_t = \beta i_t. \tag{10}$$

Deleting the "cost-push" wage shock from the wage equation, but otherwise using the price level definition and the wage-setting equation from the original two-period contract version of the Taylor model described above yields

$$p_t = .5(w_t + w_{t-1}) \tag{11}$$

$$w_t = .5(w_{t-1} + E_{t-1}(w_{t+1})) + .5g(E_{t-1}y_t + E_{t-1}y_{t+1}). \tag{12}$$

This model (6)–(12) is numerically solved for a range of parameter values. Although there are five free parameters—g, β, G, A, and μ—the behavior of output and prices has only four independent dimensions of variation. The parameters G and β affect the movement of Y_t and p_t only through their sum since substituting (10) and (9) into (8) leads to

$$y_t + p_t = (\beta + G)i_t, \tag{13}$$

and substituting (13) into (6) leads to

$$y_t = [A\{[E_t p_{t+1} - p_t] - B(p_t)\} + s_t] / (1 + AB), \tag{14}$$

where $B = (\beta + G)^{-1}$. Equation (14) reveals that the price level affects, directly and indirectly, output through two channels: there is a "destabilizing" expected price change effect and a "stabilizing" price level effect.

With this model, we pose the following question: given that the model falls short of some ideal contingent-claims Walrasian economy in many ways, does additional aggregate price flexibility—an increase in the responsiveness of wages to anticipated excess supply and demand, an increase in the coefficient g—stabilize the economy? We calculate the steady-state variance of output in response to a unit variance white-noise process for z_t, and determine whether additional price flexibility increases or decreases the steady-state variance of output for the economically relevant part of the range of parameter values.

In our initial set of numerical solutions,[3] we let A—the semielasticity of output with respect to the real interest rate—vary between 1 and 3. Assuming an average nominal interest rate of 8 percent, the .17 interest elasticity of real spending estimated by Benjamin Friedman (1978) is equivalent to an A of two. The inverse of the sum of the interest semielasticities of money demand and money supply, B, takes on three values: from .2, for an accommodating monetary policy (or a banking sector which elastically supplies real balances) and a sizable responsiveness of money demand to interest rates, to .6, representing relatively unaccommodating money supply and money demand behavior. To place these values in perspective, note that with a nominal interest rate assumed to average 8 percent and Steven Goldfeld's (1976) estimate of −.20 for the interest elasticity of money demand, completely interest inelastic money supply implies B equals .42; if the money supply has an interest elasticity of +.20, then B equals .21.

A good way to evaluate the plausibility of our parameter estimates for A and B is to calculate their implications for the simple Keynesian multiplier.[4] Holding all present

---

[3] The model is solved numerically for the steady-state variance of output by the following procedure: first, we calculate the impulse response functions for output and prices in response to a single z_t shock. To determine a unique solution, the transversaility condition that the effects of the shock eventually damp out is imposed. The model thus becomes a two-point boundary value problem which can be solved numerically along the lines of the Fair-Taylor (1983) algorithm. That is, we assume a path of expectations for prices and output after the shock, solve the model forward conditional on this path, check for convergence, update, and iterate. Second, once the impulse response function has been solved, the steady-state variance of output in response to a univariate white-noise process for z_t is easy to determine: it is simply the squared sum of the individual output responses to a unit z_t shock.

[4] But note that in the equations simulated, the multiplier has already been solved out.

TABLE 1—STEADY-STATE VARIANCE OF OUTPUT IN RESPONSE TO DEMAND SHOCKS

		$g =$						
		0	.1	.2	.4	.6	.8	1.0
$B = .2$								
$\mu = .5$	$A = 1$	93	95[a]	99[a]	110[a]	120[a]	128[a]	134[a]
	2	68	72[a]	78[a]	90[a]	100[a]	106[a]	109[a]
	3	52	56[a]	63[a]	74[a]	82[a]	86[a]	88[a]
$\mu = .75$	$A = 1$	159	163[a]	179[a]	213[a]	242[a]	267[a]	286[a]
	2	117	126[a]	148[a]	178[a]	194[a]	201[a]	202[a]
	3	89	101[a]	124[a]	157[a]	164	163	160
$B = .4$								
$\mu = .5$	$A = 1$	68	69[a]	71[a]	76[a]	81[a]	83[a]	83[a]
	2	41	42[a]	44[a]	49[a]	53[a]	56[a]	56
	3	28	22[a]	22[a]	25[a]	27[a]	28[a]	27
$\mu = .75$	$A = 1$	117	116[a]	118[a]	129[a]	140[a]	149[a]	153[a]
	2	71	70[a]	74[a]	85[a]	94	88	79
	3	47	36[a]	37[a]	42[a]	46	45	41
$B = .6$								
$\mu = .5$	$A = 1$	52	53[a]	53[a]	56[a]	59[a]	61[a]	62[a]
	2	28	28[a]	29[a]	31[a]	33[a]	34[a]	34[a]
	3	17	17[a]	18[a]	19[a]	21[a]	22[a]	22
$\mu = .75$	$A = 1$	89	87	85[a]	88[a]	93[a]	97[a]	99[a]
	2	47	46	45[a]	49[a]	54[a]	56[a]	56[a]
	3	29	28[a]	28[a]	31[a]	34	33	31

[a] Price flexibility is destabilizing at the margin.

and future expected prices constant, the simple comparative-statics multiplier obtained from equations (6)–(10), assuming a marginal propensity to spend of .67 is given by

$$M = \begin{array}{lcc} & B = .2 & B = .6 \\ \hline A = 1 & 2.5 & 2.2 \\ A = 2 & 2.1 & 1.5 \\ A = 3 & 1.9 & 1.5 \end{array}$$

An increase in autonomous spending of 1 percent of *GNP* leads in the comparative statics to an increase in nominal demand of *M* percent. These values span most estimates of this spending multiplier.

We allow g, the responsiveness of wages to excess demand or supply in the labor market, to vary between 0 and 1. Thinking of a period as one year, a value of .1 for g implies that a 1 percent reduction in output reduces the wages by .05 percent in the first year and by .1 percent after two years. In the language of Arthur Okun (1978), it implies a sacrifice ratio of 20 slightly greater than his upper-bound estimate. A value of 1 for g implies a sacrifice ratio of 2, which is somewhat below Jeffrey Sachs' (1985) optimistic reading of recent experience.

The basic numerical solutions are presented in Table 1. The degree of price flexibility increases within each row reading across. And it can be clearly seen that increasing price flexibility typically increases —not decreases—the steady-state variance of output.

Price flexibility is destabilizing at the margin in almost all cases. Only as the disturbance to the *IS* equation approaches a random walk is price flexibility ever helpful in reducing the variance of output, and then only over a relatively narrow range as the parameters move from a situation with no price flexibility to one with some price flexibility. Starting from the position of no price response to output fluctuations at all, a small amount of price flexibility can effectively damp the far-future effects of a current shock, thus reducing steady-state output variance. But very soon the margin of diminishing returns is reached because the far-future output fluctuations have already been damped, there are only small marginal returns as price flexibility is increased, and increased price flexibility disturbs the real interest rate in the present.

The pattern of responses shown in Table 1 is typical of the many calculations which we performed. In general, increases in B, A, and μ all tend to slightly increase the region over which price flexibility is stabilizing. The steeper the *LM* curve or the flatter the *IS* curve, the smaller is the scope for destabilizing price flexibility. And the more persistent are shocks, the more likely are the advantages of moving the price level quickly towards its equilibrium level to counteract the disadvantages of having a volatile rate of change of the price level.

It is interesting to note that price flexibility increases the variance of output not by increasing the persistence of shocks, but by front-loading their effects. As Figure 1 shows, the combination of price flexibility and the real interest rate effect together concentrate the output effects of a persistent nominal demand shocks in the present and near future, close in time to the moment when the shock occurs. If, in response to a monetary contraction, the price level is expected to decline substantially, then (a) real balances will be back to normal in the following period, so the effects of the contraction will not persist, and (b) the contractionary effects of high nominal interest rates will be reinforced by the extra increase in the real interest rate created by the anticipated price decline. Output is subject to shorter—and sharper in a variance metric—swings if prices are more flexible. In our forthcoming paper, we show that serial correlation in output has in fact increased over time as prices have become more and output less volatile. The serial correlation in annual deviations of output from trend has risen from approximately .4 before World War I to approximately .8 today. A common pattern—increased serial correlation and decreased volatility with more rigid prices—is present both in history and in the model. Perhaps this mechanism has something to do with the changing character of American business cycles.

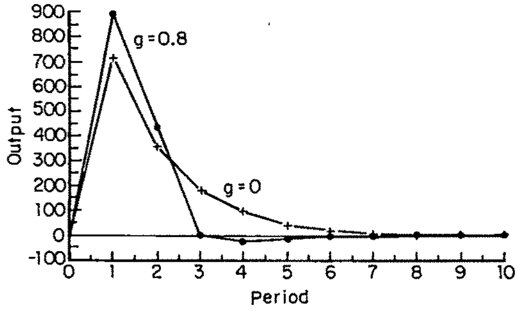


FIGURE 1. IMPULSE RESPONSE TO DEMAND SHOCK ($A=2, B=.2, \mu=.5$)

Although this model is too simple and stylized for the conclusions reached from analysis of it to be easily applied, the conclusions do seem striking. Taking Taylor's (1979) model, and modifying it so that it (i) allows for the influence of real interest rates on aggregate demand, and (ii) generates persistent output fluctuations in response to demand-side shocks, we found that this model exhibits destabilizing price flexibility as stressed by Fisher, Keynes, and Tobin for a wide range of plausible parameter values.

II. Alternative Contract Structures

A. *Changing Contract Length*

In addition to changing g as a way of modeling increases or decreases in wage and price flexibility, there is another possible type of change in price flexibility that can be analyzed with this model. It is possible to change the number and the length of the overlapping contracts. Taylor (1980) describes how to generalize his wage-setting equation from the case of two-period to the case of many-period contracts.

The previous section has shown that price flexibility is almost always destabilizing for the case of two-period contracts, but how does the variance of output change as the number of periods a given contract lasts increases from two to three to four?

Table 2 shows that, at least for the parameters chosen, increasing the length of contracts in the model frequently decreases the steady-state variance of output. We present only results for $B=.4$. With $B=.2$, increas-

TABLE 2—STEADY-STATE VARIANCE FOR INCREASING CONTRACT LENGTH
($A = 2$; $B = .4$)

		$g =$						
	Contract Length	0	.1	.2	.4	.6	.8	1.0
$\mu = .5$	2 periods	41	42[a]	44[a]	49[a]	53[a]	56[a]	57[a]
	3 periods	41	42[a]	43[a]	45[a]	47[a]	49[a]	49[a]
	4 periods	41	42	42[a]	43[a]	44[a]	45[a]	45[a]
$\mu = .75$	2 periods	71	70[a]	74[a]	85[a]	94	88	79
	3 periods	71	70[a]	72[a]	77[a]	79	78	75
	4 periods	71	70[a]	71[a]	73[a]	76	75	72

[a] Denotes that price flexibility is destabilizing at the margin.

ing contract length reduces the variance of output for all our chosen values of μ and A. As the contract length increases, the aggregate price level does not adjust as quickly to nominal shocks. And this sluggish response of the price level slightly reduces the magnitude of real interest rate fluctuations enough to reduce the variance of output in many cases.

This result casts some doubt upon the validity of arguments by those who, like Lester Thurow (1979), argue that the United States should attempt to explicitly encourage the development of "corporatist" institutions to explicitly conduct centralized wage bargaining. More generally, it calls into question standard Keynesian arguments suggesting that long-term contracts help to explain the cyclical variability of output. In the model we are working with, shortened contracts, frequent reopeners, and so forth do not necessarily stabilize output.

B. *Approaching the Limit of Perfect Markets*

There is yet another dimension of changing flexibility available which we can use to further explore the properties of this simple model: it is possible to change the "length" of the period itself, to reduce the amount of "time" taken up by each count of the index t. Instead of thinking of the model as comprising one-year periods and two-year labor contracts, we can examine the performance of the model with six-month periods and one-year contracts, or with three-month periods, and so forth. How do the effects of price flexibility change as we change the length of the period in the model?

In order to effectively halve the length of the amount of time covered by a single period in the model, the following steps are necessary: first, we set the parameter μ_{new} equal to $\sqrt{\mu_{old}}$. For the demand shock to decay in the same amount of time in the transformed model, it must decay in twice the number of periods. Second, we adjust the variance of the initial shock z_t so that the steady-state variance of the shift term s_t in the IS curve is the same as it was before.

Keeping the parameter g the same in the transformed and the untransformed model implies that the transformation to a period that covers half as much time involves a doubling of the responsiveness of the price level to output deviations.

This way of increasing price flexibility has a strong advantage over the other two possible ways—decreasing the contract length and increasing the value of the parameter g. It is not possible by varying the contract length or increasing g to create a sequence of economies which smoothly converges to the Walrasian limit. There is a tremendous difference between one- and two-period contracts, and continuously increasing g eventually leads to economically meaningless behavior on the part of the model as small expected price changes lead to enormous output jumps in the present. Shortening the amount of time covered by a single period allows for continuous movement to the Walrasian ideal, where nominal demand shocks have no real effects at all.

TABLE 3—STEADY-STATE VARIANCE APPROACHING THE WALRASIAN LIMIT ($A = 2$; $B = .4$)

	Period Length						
$g=$	1	1/2	1/4	1/8	1/16	1/32	...1/∞
0	33	33	33	33	33	33	...33
.1	33	34	34	32	25	18	...0
.2	34	35	36	33	24	16	...0
.4	36	39	41	35	24	15	...0
.6	37	43	42	36	26	15	...0
.8	38	45	43	37	27	16	...0
1.0	39	48	44	38	29	16	...0

As Table 3 illustrates, convergence toward the Walrasian limit does happen eventually: for periods of 1/8 the original or shorter, the steady-state variance of output drops relatively quickly (although by only 10–20 percent, by much less than 50 percent, with each iteration) as the period is shortened further.[5] But before the value of 1/8, convergence toward the limit is not, or is at most only slightly, visible. The effect of a shortened period in causing more rapid price changes—and thus more of an incentive to postpone or accelerate spending by one period—approximately balances the stabilizing effects of more rapid price flexibility. In the limit, increased price flexibility is indeed stabilizing. But one must already be very close to that Walrasian limit before its properties become a good guide to the behavior of the model.

III. Supply Shocks

Although increased price flexibility clearly often is destabilizing at the margin in response to demand shocks in our models, increased price flexibility is—as Driskill and Sheffrin have proved—stabilizing at the margin in response to supply-side shocks to the level of wages. This raises an interesting question: if there is a mixed distribution of shocks, how heavily does the mix have to be tilted towards supply shocks before there ceases to be a region where increased price flexibility is destabilizing at the margin? If we change our wage-determination equation to add a supply shock:

$$(15)\quad W_t = .5(W_{t-1} + E_{t-1}W_{t+1}) + x(.5E_{t-1}(Y_t + Y_{t+1})) + N_t,$$

how large does σ_N^2 have to be relative to σ_z^2 for the region where price flexibility is destabilizing at the margin to disappear?

In our model the answer turns out to be somewhat surprising. As price flexibility increases, the variance of output generated by a serially uncorrelated supply shock quickly declines and approaches an asymptote for the reasonable parameter values used here, as Table 4 reports.

There is thus a region where (*i*) further increases in price flexibility do little to reduce output variance generated by supply-side shocks, and (*ii*) further increases still add to the variance generated by demand shocks. Therefore the distribution of shocks must be heavily weighted toward the supply side for the region where price flexibility is destabilizing to disappear.

Table 5 reports some sample output variances for different distributions of independent shocks. The variance of N_t, the supply shock, is taken to be one; the variance of the demand shock, z_t, is taken to be B.

The intuition behind this result is rather simple. Driskill and Sheffrin show that the output response of a real interest rate-enhanced Taylor-like model to a supply shock is an ARMA(11) process. As the degree of price flexibility increases, the autoregressive term becomes smaller and smaller because the economy moves to eliminate effects that are not due to this period's and last period's surprise shock. In the limit, state variables become simple one-period moving averages of shocks.

Now the autoregressive parameter turns out typically to be already quite small for relatively low g. It is not uncommon for an autoregressive parameter of .3 to be associ-

---

[5] Results are presented for $B = .4$. Results for $B = .2$ show even slower convergence toward the Walrasian limit.

TABLE 4—STEADY-STATE VARIANCE IN RESPONSE TO SUPPLY SHOCKS

		$g=$						
	$A=$	0	.1	.2	.4	.6	.8	Asymptote
$B=.2$	1	∞	30	28	27	27	26	26
	2	∞	85	80	78	76	76	75
	3	∞	142	136	132	130	130	130
$B=.4$	1	∞	37	33	31	29	28	27
	2	∞	84	76	71	68	66	65
	3	∞	123	112	105	102	100	99

TABLE 5—STEADY-STATE VARIANCE IN RESPONSE TO MIXED SHOCKS
($A=2$; $\mu=.5$)

	Ratio of Supply to Demand Shock Variance	$g=$						
		0	.1	.2	.4	.6	.8	1.0
$B=.2$	1	∞	157[a]	158[a]	168[a]	177[a]	181[a]	183[a]
	2	∞	121	119[a]	123[a]	127[a]	130[a]	131[a]
	5	∞	99	96	95[a]	96[a]	97[a]	97[a]
	10	∞	92	88	86[a]	86[a]	87[a]	87
$B=.4$	1	∞	126	120	120[a]	121[a]	122[a]	122[a]
	2	∞	105	98	95	96[a]	97[a]	97[a]
	5	∞	92	85	81	79	79[a]	79
	10	∞	88	80	76	74	74	72

[a] Denotes that price flexibility is destabilizing on the margin.

ated with a g of similar magnitude. The contribution of the "out-period" terms of an ARMA(1,1) to the total variance of the process is approximately the square of the autoregressive parameter times the contribution of the first two "in-period" terms, in which shocks enter the moving average portion of the process. Therefore, even with a g of only 0.3, the variance caused by supply shocks is already only some 9 percent above its asymptotic value. There is no room for further increases in price flexibility to produce significant declines in supply-shock-driven output variability.

We conclude that the introduction of supply shocks into our simple model is unlikely to remove our central result: that there is a significant and reasonable region of parameters for which price flexibility is destabilizing. Even if shocks on the supply side play a quantitatively significant role in cyclical output fluctuations, additional price flexibility may well do relatively little to damp their effects and may amplify the output fluctuations induced by demand shocks.[6]

IV. Alternative Determinants of Aggregate Demand

The previous sections have all assumed that what matters for the determination of aggregate demand is the short-term real in-

---

[6] Taylor (1980) concludes that the phenomenon of destabilizing price flexibility is not of empirical significance. He relies on the relatively primitive model developed in our forthcoming article with inertial expectations in the price equation, that lacks any attempt at explicit microfoundations, and generates output movements in response to demand shocks that have negative serial correlation. Taylor finds that for the primitive model with supply and demand shocks, there are regions where price flexibility is stabilizing and regions where price flexibility is destabilizing, but that empirically reasonable parameter values are likely to be contained in the stabilizing region. We find it hard to judge what parameter values are reasonable in such a stylized model.

TABLE 6—STEADY-STATE VARIANCE FOR A Q THEORY OF AGGREGATE DEMAND ($\mu = .5$; $\delta = .20$)

		$g =$					
	$A =$	.1	.2	.4	.6	.8	1.0
$B = .2$	1	78.0^a	79.0^a	79.6^a	80.4^a	78.4	75.1
	2	52.0^a	53.6^a	55.8^a	54.3	54.1	50.8
	3	37.3^a	39.1^a	39.8^a	40.8	40.4	37.2
$B = .4$	1	53.0	52.9^a	53.1^a	53.4^a	52.5	50.1
	2	27.9^a	28.4^a	29.1^a	28.9	28.8	27.4
	3	17.1^a	17.6^a	18.2^a	18.1	17.9	16.1
$B = .6$	1	39.4	39.1	39.0	39.2^a	38.7	37.5
	2	17.6^a	17.8^a	18.0^a	18.1	18.0	17.6
	3	9.7^a	10.0^a	10.2^a	10.2	10.1	9.9

[a] Denotes that price flexibility is destabilizing at the margin.

terest rate, the one-period interest rate r_t. But there is an important and convincing line of thought which argues that the relevant interest rate for the purpose of determining investment is a long-term interest rate. Suppose that the level of aggregate demand is determined by the level of investment, which is in turn determined by Tobin's q. How does this shift in the mechanism that determines aggregate demand change the results of the previous sections?

Blanchard and Sachs (1982) consider a model with quantity-constrained demands, fully optimizing agents, and durable capital that is costly to adjust. They conclude that the Mundell effect is not strong enough to be destabilizing. In their model, price flexibility amplifies fluctuations in the short-term real interest rate. But because price flexibility leads to the faster damping of disturbances, long-term rates are less erratic and output is less variable.

A critical issue here is the length of time which constitutes the "long run," the horizon with which Tobin's q is calculated. In the spirit of the previous models in this paper, we replace r_t in (8) with its closest analogue in a q-theoretic framework, $q_t - 1$:

$$(16) \qquad (1 + AB)y_t = A(q_t - 1) + s_t.$$

Arbitrage implies that q_t evolves according to

$$(17) \qquad q_t = (1 - K - r_t)q_{t+1} + (F_k)_t,$$

where $(F_k)_t$ is the return paid on a unit of capital in place in period t, K is the sum of the risk premium on equity and rate of depreciation, and r_t is the real short-term interest rate. Clearly the larger is K, the closer is q_t to $1 - r_t$. If K is "high enough" in some sense, price flexibility will be destabilizing even in this q-theoretic framework. The process determining q_t will place very high weight on the more volatile real return fluctuations in the present and less weight on the more stable real returns in the far future. The interesting question is whether K is likely to be high enough to render price flexibility destabilizing in actual economies.

Since World War II, the real return on equities in the United States has averaged some nine or so percentage points higher per year than the return on Treasury bills. One component of K is therefore this risk premium. It is reasonable to assume that capital depreciates at 10 percent per year. We therefore assume that $K = 20$ in the calculations reported below, although our results are somewhat sensitive on this assumption.[7]

The model used for the simulations reported in Table 6 is our basic model with two-period contracts, only with the IS aggregate demand equation (6) replaced by (16), and with the arbitrage equation (17) included with a value of .20 for K. Profits $(F_k)_t$ are assumed constant; allowing for procyclical profits introduces an accelerator into the model without changing its qualitative behavior. Results are presented for $\mu = 5$. Price flexibility is more destabilizing for lower values, and less destabilizing for higher values of μ.

---

[7] If K is lower, then the destabilizing effect of price flexibility is reduced. With $K = 16$, destabilizing effects are severely damped. And if K is lowered to 10, then price flexibility is never destabilizing for the parameter values considered.

Price flexibility is still destabilizing over a range, although not over as large a range as the basic model. Moving to a theory of aggregate demand based on the long-term rather than on the short-term interest rate weakens but does not eliminate the plausibility of destabilizing price flexibility. With higher values of B, the traditional conclusion that price flexibility is stabilizing is restored.

As Robert Hall (1977) emphasizes, there are strong reasons to expect investment to depend on short as well as long real rates. Timing is a critical aspect of many investment decisions. For many types of capital goods, adjustment costs are not important. And so the financial variable that should enter into the determinant of aggregate demand should be neither r_t nor q_{t-1}, but instead some weighted average.

There are several other reasons to think that the weight placed on r_t should be relatively high. Much cyclically sensitive capital formation is nondurable, or not very durable. Moreover, there may be effects which are well-correlated with the short-term interest rate which are not in representative decision-maker models. Agents may be limited in their spending by liquidity constraints. And sharp declines in the price level that were unanticipated at the moment many contracts specifying financial liability were written can have large effects on aggregate demand through the distributions of wealth. Putting a high weight on r_t would lead us back towards the results in earlier sections.

Destabilizing price flexibility is not removed as a possibility by the move from using short-term to using long-term interest rates in the determination of aggregate demand. Our initial conclusion that models derived from Taylor (1979, 1980) can exhibit destabilizing price flexibility for a wide range of plausible parameter values has turned out to be quite robust. If Taylor's models can serve as a reasonable base from which to reason about the behavior of the economy, the results of this paper show that there is every reason to take Keynes seriously when he challenges the view that increased price flexibility would reduce the seriousness of business cycles.

V. Conclusions

Our results suggest that standard Keynesian models wage and price rigidity have the implication that increases in price flexibility starting from degrees of flexibility like those observed in economies such as the United States may well increase the cyclical variability of output. This conclusion holds even allowing for forward-looking behavior in wage-setting and in financial markets. This observation makes it unlikely that the extra degree of nominal rigidity of the type associated with union contracts, for example, can plausibly be blamed for cyclical fluctuations in output. Even if increased nominal rigidity is not stabilizing, it is unlikely to be significantly destabilizing. As we have emphasized in our forthcoming paper, these observations are consistent with the broad sweep of American macroeconomic history. The increasingly non-Walrasian character of the economy and the associated reductions in wage and price flexibility have coincided with improvements in macroeconomic performance.

The mainline American Keynesian view that the major difference between economies that are well-approximated by classical models and economies that are well-approximated by Keynesian models lies in the fact that the first set of economies have more aggregate price flexibility is particularly odd in view of the origins of Keynesian economics in the experience of the Great Depression. Between 1929 and 1932, the U.S. price level declined by 9 percent a year. In 1932, U.S. real *GNP* was two-thirds of what it had been three years earlier. It is difficult to believe that the problems of the U.S. economy during the Great Contraction were due to the fact that prices did not adjust quickly enough to absorb the nominal shock which came either from inappropriate monetary policy or from pessimistic animal spirits. It seems implausible to assert that if only the price level had declined by 20 instead of 9 percent per year, that the U.S. economy would have had a high level of output in 1932. Rather more plausible is the belief that if the price level had fallen at 20 percent per year, the contraction would have been even more serious as very high real interest rates would have drastically reduced the level of economic activity.

Events surrounding the trough of the Great Depression lend some support to our position, although a convincing analysis must wait for the future. During the decline of 1929–33, the view that price declines were a cause—as well as a consequence—of depressed output gathered force. President Roosevelt was greeted on his inauguration by demands for a rollback of deflation, demands which found expression in the National Industrial Recovery Act. According to Ellis Hawley, American business leaders believed that "excessive competition," by depressing wages and product prices without writing down fixed debt-related costs, "had destroyed profitable operation, undermined business confidence, and reduced the rate of investment" (1966, p. 27). According to W. Arthur Lewis, the NIRA was intended to "promote recovery... by raising prices, and thus dissipating the gloom of business men... and by raising wages" (1949, p. 107).

The NIRA was signed in June 1933, a few months after the trough in both output and prices. Over the period of its operation, it certainly raised prices. Output also recovered somewhat, and clearly the recovery was not triggered by any further decline in nominal interest rates or by fiscal policy (as E. Cary Brown, 1956, has convincingly demonstrated). In Lewis' judgment, the initial "psychological effect [of the NIRA] was good; business men took heart from the prospect of rising prices" (p. 107). Nevertheless the possible stimulative effects of the NIRA operating through price expectations remain speculative. Rolling back deflation was only one of the policies embedded in the NIRA. The NIRA's encouragement of cartelization may have had contractionary macroeconomic effects of its own.

International evidence corroborates our skepticism about the causal role of substantial nominal rigidities in explaining business cycles. In much of Latin America, high rates of inflation are endemic. Perhaps because of the increased wage and price flexibility that accompanies rapid inflation, real interest rates are much more variable than in the United States and so too is real output. Despite a high degree of price flexibility, monetary and fiscal policies have potent effects on the level of economic activity.

As Franklin Fisher (1984) emphasizes, economics does not possess a fully satisfactory theory of price adjustment. There does not yet exist a convincing demonstration that a decentralized economy will rapidly converge to its Walrasian equilibrium after being shocked. The stability of an economy without institutional impediments to price flexibility in the face of demand shocks is an assumption, not a conclusion of economic analysis. The performance of actual economies suggests that it may not be a particularly good one.

While calling into question the common conclusion that more nominal rigidities mean more variability in output and employment, our results support arguments suggesting the importance of the existence of nominal rigidities in explaining cyclical fluctuations. A problem for "menu" arguments explaining business cycles is that the rigidities explained seem small relative to the length and depth of the fluctuations that must be explained. Our analysis suggests that because of the effects of expected price changes, even small amounts of nominal rigidity can be substantially destabilizing.

REFERENCES

Akerlof, George and Yellen, Janet, "A Near-Rational Model of the Business Cycle," *Quarterly Journal of Economics*, Suppl. 1985, *100*, 823–37.

Blanchard, Olivier J., "Menu Costs, Monopolistic Competition, and the Real Effects of Nominal Money," xerox, 1985.

_____ **and Sachs, Jeffrey,** "Anticipations, Recessions and Policy," *Annales de l'INSEE*, 1982, *47–48*, 117–44.

_____ **and Wyplosz, Charles,** "An Empirical Structural Model of Aggregate Demand," *Journal of Monetary Economics*, January 1981, *7*, 1–28.

Brown, E. Carey, "Fiscal Policy in the "Thirties: A Reappraisal," *American Economic Review*, December 1956, *46*, 857–79.

Cagan, Phillip, "Changes in the Recession Behavior of Wholesale Prices," *Explorations in Economic Research*, 1975, *2*, 54–103.

De Long, J. Bradford and Summers, Lawrence, "The Changing Cyclical Variability of Economic Activity in the United States," in R. J. Gordon, ed., *The American Business Cycle: Continuity and Change*, Chicago: University of Chicago Press, forthcoming.

Driskill, Robert and Sheffrin, Steven, "Is Price Flexibility Destabilizing?" *American Economic Review*, September 1986, *76*, 802–07.

Fair, Ray and Taylor, John, "Solution and Maximum Likelihood Estimation of Dynamic Nonlinear Rational Expectations Models," *Econometrica*, July 1983, *51*, 1169–86.

Fischer, Stanley, "Contracts, Credibility and Disinflation," NBER Working Paper No. 1339, 1984.

Fisher, Franklin, *Disequilibrium Foundations of Equilibrium Economics*, Cambridge: Cambridge University Press, 1984.

Fisher, Irving, "The Business Cycle Largely a 'Dance of the Dollar'," *Journal of the American Statistical Association*, December 1923, *18*, 1024–28.

______, "Our Unstable Dollar and the So-Called Business Cycle," *Journal of the American Statistical Association*, June 1925, *20*, 179–202.

Friedman, Benjamin, "Crowding Out or Crowding In? Economic Consequences of Financing Government Deficits," *Brookings Papers on Economic Activity*, 3:1978, 593–641.

Goldfeld, Steven, "The Case of the Missing Money," *Brookings Papers on Economic Activity*, 3:1976, 683–739.

Gordon, Robert J., *Milton Friedman's Monetary Framework: A Debate with His Critics*, Chicago: University of Chicago Press, 1974.

______, "A Consistent Characterization of a Near-Century of Price Behavior," *American Economic Review Proceedings*; May 1980, *70*, 243–49.

Hall, Robert E., "Investment, Interest Rates, and the Effects of Stabilization Policies," *Brookings Papers on Economic Activity*, 1:1977, 61–120.

Hawley, Ellis W., *The New Deal and the Problem of Monopoly*, Princeton: Princeton University Press, 1966.

Jonung, Lars, "Comparative Responses to the Great Depression: Sweden and the U.S.," in Karl Brunner, ed., *The Great Depression Revisited*, Boston: Kluwer-Nijhoff, 1981.

Keynes, John Maynard, *The General Theory of Employment, Interest and Money*, London: Macmillan, 1936.

Lewis, W. Arthur, *Economic Survey*, London: George Allen and Unwin, 1949 .

McCallum, Bennett T., "The Liquidity Trap and the Pigou Effect: A Dynamic Analysis with Rational Expectations," *Economica*, November 1983, *50*, 395–405.

Mankiw, N. Gregory, "Small Menu Costs and Large Business Cycles," *Quarterly Journal of Economics*, May 1985, *100*, 529–37.

Okun, Arthur, "Efficient Deflationary Policies," *American Economic Review Proceedings*, May 1978, *68*, 348–52.

______, *Prices and Quantities*, Washington: The Brookings Institution, 1981.

Sachs, Jeffrey, "The Changing Cyclical Behavior of Wages and Prices, 1890–1976," *American Economic Review*, March 1980, *70*, 78–90.

______, "The Policy Mix in 1985," *Brookings Papers on Economic Activity*, 1:1985, 117–83.

Schultze, Charles, "Some Macro Foundations for Micro Theory," *Brookings Papers on Economic Activity*, 2:1981, 521–91.

Taylor, John, "Staggered Wage Setting in a Macro Model," in *American Economic Review Proceedings*, May 1979, *69*, 108–113.

______, "Aggregate Dynamics and Staggered Contracts," *Journal of Political Economy*, February 1980, *88*, 1–23.

Thurow, Lester, *The Zero-Sum Society*, New York: Basic Books, 1979.

Tobin, James, "Keynesian Models of Recession and Depression," *American Economic Review Proceedings*, May 1975, *65*, 195–202.

Efficient Contracts with Costly Adjustment: Short-Run Employment Determination for Airline Mechanics

By DAVID CARD*

This paper presents an empirical analysis of firm-specific employment and wage outcomes for mechanics in the domestic airline industry. A dynamic contracting model is presented that incorporates both costly employment adjustment and potential gaps between contract wage rates and the opportunity value of workers' time. The model gives a useful description of the employment-output linkage in the data, but is less successful in capturing the dynamic relation between employment, contract wage rates, and wage rates outside the airline industry.

The search for a credible interpretation of observed employment patterns has led to widespread interest in the notion of efficient labor contracts.[1] In contrast to traditional auction models of the labor market, which relate fluctuations in employment directly to changes in wage rates, contracting models permit a more flexible link between wage payments and employment determination. According to the simplest version of the efficient contracting hypothesis, in fact, the level of employment maximizes the joint income of workers and the firm, while the level of wages represents a pure transfer between them. If this "strong form" efficiency hypothesis is correct, then it calls into question a wide variety of policy conclusions based on the assumed distortionary effects of union wage differentials and short-term wage rigidities.[2]

A simple test of the efficient contracting hypothesis is provided by a firm-level employment equation that includes both the wage rate at the firm and some measure of the alternative wage available to workers.[3] According to the strong form efficiency hypothesis, the joint income of workers and the firm is maximized when the marginal value product of labor equals its outside opportunity wage. If this hypothesis is correct, employment is independent of firm-specific wage rates and depends only on the alternative wage rate. In the traditional labor demand model, by comparison, the firm's profits are maximized when the marginal value product of labor equals the firm-specific wage. If this model is correct, then employment is independent of alternative wage rates and depends only on the firm's wage rate. Finally, in a general contracting model, labor is allocated between outside opportunities and contract employment on the basis of a shadow value that varies with both firm-specific and alternative wage rates.

---

*Department of Economics, Princeton University, Princeton, NJ 08544. I am grateful to Robert LaLonde for his assistance in gathering the data in this paper, and to Jim Brown, John Pencavel, and an anonymous referee for comments on earlier drafts.

[1] The theoretical literature on efficient contracting is voluminous: see in particular Wassily Leontief (1946), Costas Azariadis (1975), Martin Baily (1974), Donald Gordon (1974), Robert Hall and David Lilien (1979), Ian McDonald and Robert Solow (1981), and the recent surveys by Oliver Hart (1983) and Sherwin Rosen (1985). Jan Svejnar (1986) presents an empirical study of efficient wage and employment outcomes. James Brown and Orley Ashenfelter (1986), Thomas MaCurdy and John Pencavel (1986), and F. Martinello (1984) attempt to test between efficient contracting models and conventional employment-setting models.

[2] This point was made forcefully by Robert Barro (1977) in a comment on models of nominal wage contracting.

[3] This is essentially the test procedure adopted by Brown and Ashenfelter. MaCurdy and Pencavel estimate a capital-labor ratio equation that includes both contract and alternative wage rates.

In a dynamic setting, however, tests of even simple efficient contract models are complicated by the fact that alternative wage rates may help to predict future contract wage rates. Suppose, for example, that employment adjusts toward a target level that depends on a weighted average of future contract wage rates. In this case, even if desired employment is independent of alternative wages, current and lagged alternative wage rates may enter the employment equation as predictors of future contract wages.

In a dynamic model it is therefore necessary to sort out two competing routes for the alternative wage to influence the level of contract employment: (*i*) directly, through the appropriate expression for the shadow value of labor; and (*ii*) indirectly, through the statistical link between current alternative wage rates and expected future contract wage rates.

This paper presents an empirical analysis of the potential links between wage rates and employment for mechanics in the domestic airline industry. The data consist of quarterly observations on employment, output, and wages for seven major airlines, drawn from the period prior to deregulation of the industry. Wage rate information is taken from union contracts covering mechanics at each of the seven airlines, while employment and output data are taken from Civil Aeronautics Board records.

Section I presents a preliminary analysis of the data using unrestricted vector autoregressions. The analysis reveals a striking similarity between the serial correlation properties of these micro-level data, and the properties of more familiar aggregate data. The analysis also shows that aggregate manufacturing wage rates are an important determinant of both employment levels and wage rates for airline mechanics. On the assumption that the manufacturing wage rate represents the alternative wage rate for airline mechanics, the observed link between mechanics' employment and manufacturing wages can be attributed either to efficient contracting considerations, or to the fact that manufacturing wages help predict future contract wage rates, or both.

The second section of the paper presents a simple intertemporal contracting model in which it is possible to disentangle these two effects. The model assumes that employment and wages are selected to minimize the cost of aircraft maintenance, subject to a utility constraint for mechanics. Adjustment costs are introduced on the firm's side to generate an employment function with serial persistence and gradual adjustment to output shocks. Two specifications of workers' preferences are presented that yield alternative expressions for the shadow value of labor in an optimal contract. The traditional labor demand model and the strong form efficient contracting model are obtained as special cases of the general model.

The model yields a partial-adjustment employment equation that expresses current employment in terms of lagged employment, lagged departures, and lagged wage rates. The model generates testable restrictions across the employment equation and the forecasting equations for output and wages, and summarizes the employment effects of alternative wage rates in terms of two components: a direct effect on the desired level of employment; and on indirect effect on forecasts of future contract wages.

Section III presents the results of fitting the model to data on departures, wages, and employment for the seven airline firms. The empirical analysis is generally unsupportive of either the strong form efficient contracting model or the labor demand model. Both models are rejected in favor of a more general contracting model in which the opportunity cost of employment is a weighted average of the contract wage and the alternative wage. The parameter estimates for all three models are poorly determined, however, and the implied reduced forms fail to reconcile all the dynamic linkages between contract wages, alternative wages, and employment.

I. Preliminary Data Analysis

The data in this paper consist of quarterly observations on employment and wages for aircraft mechanics employed in the domestic operations of seven airlines: American, Braniff, Continental, Eastern, Trans World, United, and Western.[4] Employment, earnings, and flight data were collected from various Civil Aeronautics Board (CAB) sources for the period 1969:I to 1976:IV.[5] Wage rates were collected from union contracts summarized in *Current Wage Developments* and the Bureau of National Affairs' *Daily Labor Report*.[6] A complete description of data sources is provided in Appendix II.

Mechanics at these airlines are represented by three unions: the Transport Workers Union (TWU) at American; the Teamsters (IBT) at Western; and the Machinists (IAM) at the other five airlines. Although the IAM bargained individually with these five airlines during the sample period, contract terms and expiration dates differed little between them. Differences among the IAM contracts were due mainly to delays in signing new contracts at the individual airlines.[7]

Table 1 presents annual data on nominal and real wage rates for airline mechanics at the various firms between 1969 and 1976.[8] There were only small and unsystematic differences between wage rates negotiated by the three unions. Compared to other workers, however, airline mechanics earned relatively high wage rates during this period: 50 to 60 percent higher than average straight-time hourly earnings in manufacturing, for example, and about 25 percent higher than average hourly earnings reported by maintenance mechanics in manufacturing. Airline mechanics also earned a small premium (5 to 15 percent) over unionized mechanics at the major aircraft companies.

Airline mechanics' nominal wage rates are established in two- or three-year contracts that typically include both noncontingent deferred increases and cost-of-living allowance clauses. During the sample period, new agreements were negotiated in 1969 (from March to December, depending on airline), 1971–72 (May 1971 to December 1972), 1973–74 (November 1973 to August 1974), and 1975–76 (December 1975 to September 1976). Because of deferred increases and cost-of-living adjustments, however, the relation between contract negotiations and real wage rates is indirect. Real wage rates increased over the term of the 1969 contracts, for example, but were more or less constant between 1973 and 1976.

The behavior of real contract wage rates is analyzed more formally in Table 2. Column 1 presents a simple second-order autoregression (AR(2)) fit jointly to the logarithms of contract wage rates for all seven airlines, with unrestricted constants, trends and seasonal dummy variables for each airline.[9] As is apparently true for aggregate real wage

---

[4] Data are also available for three other domestic trunk airlines—Delta, National, and Northwest—as well as for Pan Am. Because of the high level of strike activity at National and Northwest, these airlines were excluded from the present study. Pan Am differs from the domestic trunks in that a large share of its business is international. For this reason, domestic employment data for Pan Am may be misleading, and I chose to exclude it. Delta's mechanics are nonunionized and, as a result, no direct measure of contract wage rates is available.

[5] Detailed quarterly employment data are unavailable after 1976. Employment and payroll information pertain to maintenance and related workers, as defined by the CAB. This group includes aircraft inspectors and mechanics, as well as cleaners, janitors, and stock clerks. According to CAB records, there were 37,036 maintenance workers in the domestic trunk airlines in the third quarter of 1975. A Bureau of Labor Statistics (BLS) *Industry Wage Survey* during August–November 1975 counted 29,518 inspectors and mechanics, and 8,588 cleaners, janitors and stock clerks at the domestic trunks, for a total of 38,106 maintenance and related workers. On this basis, approximately 80 percent of maintenance workers are actually airline mechanics or inspectors.

[6] Wage rates for mechanics at Western Airlines were obtained from copies of the contracts generously made available to me by the IBT Airline Division.

[7] During this period the delay between expiration of old contracts and renegotiation of new contracts was typically six to twelve months.

[8] Wage rates in Table 1 are for certified mechanics, excluding premiums for line service work (10 to 25 cents per hour during this period) and FAA licenses (10 to 20 cents per hour per license). Wage rates for mechanics represented by the IAM are summarized by the wage rates at United Airlines.

[9] A simple F-test that the AR(2) coefficients in the wage equation are the same across all seven airlines yields a probability value of .04. The major difference in coefficients is between American and the other airlines.

TABLE 1—WAGE RATES FOR AIRLINE MECHANICS AND COMPARISON GROUPS: 1969–76

	American–TWU[a]	United–IAM[b]	Western–IBT[c]	Maintenance Mechanics in Manufacturing[d]	Production Workers in Manufacturing[e]	Licensed Mechanics Boeing–IAM[f]
1969	4.60	4.40	4.47	3.77	3.07	4.46
	[4.60]	[4.40]	[4.47]	[3.77]	[3.07]	[4.46]
1970	5.10	5.20	5.27	3.98	3.24	4.65
	[4.81]	[4.91]	[4.98]	[3.76]	[3.06]	[4.39]
1971	5.60	5.60	5.68	4.24	3.44	4.90
	[5.07]	[5.07]	[5.14]	[3.84]	[3.11]	[4.44]
1972	6.20	5.95	5.98	4.56	3.65	5.38
	[5.43]	[5.21]	[5.24]	[4.00]	[3.20]	[4.71]
1973	6.40	6.60	6.45	4.84	3.90	5.78
	[5.28]	[5.44]	[5.32]	[3.99]	[3.22]	[4.77]
1974	6.85	7.10	6.85	5.22	4.23	6.16
	[5.09]	[5.28]	[5.09]	[3.88]	[3.14]	[4.58]
1975	7.50	7.70	7.70	5.97	4.65	7.03
	[5.10]	[5.24]	[5.24]	[4.07]	[3.17]	[4.79]
1976	8.10	8.45	7.90	6.54	5.00	7.53
	[5.22]	[5.44]	[5.09]	[4.22]	[3.22]	[4.85]

Notes: Real wage rates represent nominal rates divided by the CPI, all items, 1969 = 100. In each column, real wage rates are shown in brackets below nominal rates.

[a] Average base rate in effect for mechanics in collective agreement between American Airlines and Transport Workers Union.

[b] Average base rate in effect for mechanics in collective agreement between United Airlines and Machinists Union. Wages for mechanics at Braniff, Continental, Eastern, TWA, and United are similar.

[c] Average base rate in effect for mechanics in collective agreement between Western Airlines and Teamsters Union.

[d] Average hourly earnings of maintenance mechanics in manufacturing, from the BLS Area Wage Survey.

[e] Average straight-time hourly earnings of production workers in manufacturing, annual averages of quarterly values.

[f] Average base rate in effect for licensed aircraft mechanics in collective agreement between Boeing and Machinists Union from the BLS Wage Chronology.

rates, the contract wage for airline mechanics is approximately a first-order autoregressive process, with something less than a unit autoregressive coefficient.[10] For purposes of comparison with the results from a longer sample period, column 6 of the table presents the same representation of contract wage rates fit to an extended sample period (1964 to 1978). The addition of seven extra years of data has no appreciable impact on the estimated coefficients, however.

Columns 2, 3, and 4 of Table 2 contain estimates of real wage regressions that include lagged values of manufacturing wages and consumer prices. Each of these two aggregate series has strong causal links to contract wages. The point estimates in column 2 show that a permanent 1 percent increase in manufacturing wages leads to an eventual increase in contract wages on the order of 1 percent, while the estimates in column 3 suggest that a permanent increase in prices leads to a permanent decrease in contract wages. In column 4, both prices and manufacturing wages are entered in the regression for contract wages. Although the statistical significance of the individual coefficients is mixed, a test that lagged consumer prices improve the forecast of contract wages, given lagged manufacturing wages, is marginally

---

[10] See, for example, Ashenfelter's and my paper (1982). A similar specification fit to real straight-time average hourly earnings in manufacturing over this sample period yields a first-order coefficient of 1.09 (with a standard error (*SE*) of .20) and a second-order coefficient of −.26 (*SE* = .20).

TABLE 2—AUTOREGRESSIVE REPRESENTATIONS OF REAL WAGE RATES[a]

	Contract Real Wage: 1969:III–1976:IV					Contract Real Wage: 1964:III–1978:IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Contract Wage	.71	.61	.63	.60	.61	.69	.66	.66	.66
$(t-1)$	(.08)	(.07)	(.08)	(.07)	(.07)	(.05)	(.05)	(.05)	(.05)
Contract Wage	−.04	−.04	−.12	−.09	−.12	−.02	−.01	−.03	−.01
$(t-2)$	(.07)	(.07)	(.07)	(.07)	(.07)	(.05)	(.05)	(.05)	(.05)
Manufacturing	–	2.04	–	.74	.86	–	1.48	–	.70
Wage $(t-1)$		(.39)		(.57)	(.52)		(.36)		(.47)
Manufacturing	–	−1.37	–	−1.08	−1.19	–	−.95	–	−.62
Wage $(t-2)$		(.40)		(.52)	(.47)		(.36)		(.44)
Price Index	–	–	−1.86	−1.72	−1.56	–	–	−1.62	−1.14
$(t-1)$			(.42)	(.64)	(.59)			(.40)	(.54)
Price Index	–	–	1.45	1.13	.94	–	–	1.40	.96
$(t-2)$			(.49)	(.53)	(.50)			(.39)	(.51)
Employment	–	–	–	–	−.07	–	–	–	–
$(t-1)$					(.06)				
Employment	–	–	–	–	.12	–	–	–	–
$(t-2)$					(.06)				
Probability Value for Exclusion of:									
Manufacturing Wage	–	.00	–	.15	.06	–	.00	–	.15
Consumer Prices	–	–	.00	.04	.03	–	–	.00	.05

Notes: Regressions include constants, linear trends, and quarterly dummy variables. Coefficients (except constants, trends, and seasonals) are restricted to be equal across airlines. Contract wages and manufacturing wages are deflated by the Consumer Price Index. Regressions contain airline-specific dummy variables for strike and immediate poststrike observations.

[a] The standard errors are shown in parentheses.

significant, while a test for lagged manufacturing wages, given lagged prices, is not quite significant at conventional levels. Parallel results for the longer sample period (in cols. 7, 8, and 9) lead to very similar conclusions.

Finally, column 5 reports the coefficients of lagged employment in forecasting contract wages. The point estimates suggest that increases in employment lead to small increases in real contract wages, although a test of the hypothesis that employment fails to Granger-cause wages is not significant at conventional levels.[11]

The general pattern of employment for airline mechanics is illustrated in Figure 1, which gives a time-series plot of departures and employment at American Airlines. In addition to mechanics' employment, the figure shows quarterly employment levels for pilots and flight attendants. One of the most interesting features of the data is the extent to which employment of these three groups of workers is smoothed vis-à-vis departures. At one extreme, the number of employed pilots is very stable and shows little relation to output. At the other, the number of flight attendants is quite variable over time. Mechanics fall somewhere between these two groups. Fluctuations in departures translate into dampened fluctuations in the number of employed mechanics.

Data on flying operations and mechanics' employment are summarized by airline in Table 3. The airlines fall naturally into two groups: the four largest firms—American, Eastern, Trans World, and United—and the three smaller firms. Interestingly, the number of mechanics per unit of output is quite different between the larger and smaller airlines. Relative to the employment levels at the smaller airlines, the larger airlines have 60 to 70 percent more maintenance employees per departure (or per available seat mile). To some extent this may reflect a greater reliance on outside contractors to

---

[11] The probability value of the test statistic is 0.12.

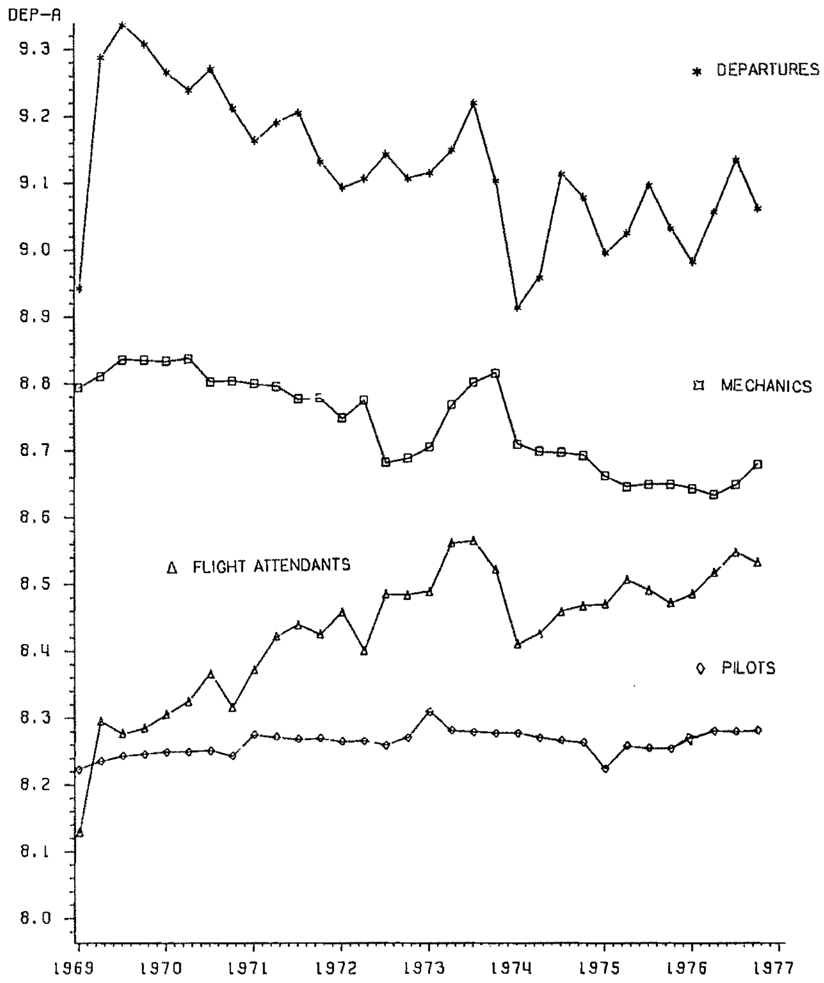


FIGURE 1. EMPLOYMENT AND DEPARTURES: AMERICAN AIRLINES

perform specialized maintenance operations at the smaller airlines. Smoothing of employment relative to departures is indicated by the smaller coefficient of variation of employment for all the airlines except Continental and United. A similar conclusion emerges from a comparison between detrended and deseasonalized employment and departures. With the exception of United Airlines, the standard deviation of the logarithm of employment is about two-thirds as large as the standard deviation of the logarithm of departures, when both series have been fitted to a linear trend and quarterly dummy variables.

Table 4 presents a more formal analysis of employment variability and its relation to departures and wages. Column 1 presents a simple AR(2) specification of employment, fit jointly to the seven airlines with firm-specific trends, constants, and seasonals. Employment of airline mechanics displays the hump-shaped moving average representation that characterizes many aggregate employment time-series.[12] The response to a unit innovation in employment persists for

---

[12] See Thomas Sargent (1978), for example.

TABLE 3—MEANS AND STANDARD DEVIATIONS: DOMESTIC AIRLINE DATA 1969:I–1976:IV

	American	Braniff	Continental	Eastern	TWA	United	Western
Average Quarterly Departures	92980	41230	35250	119570	70210	140000	37380
Standard Deviation of Departures	9740	2870	2370	820	4420	14720	1860
Average Quarterly Maintenance Employment	6255	1072	1109	6209	6358	9170	1336
Standard Deviation of Employment	440	57	95	189	223	1002	47
Average Number of Seats per Flight	132	116	129	107	126	134	121
Average Flight Length	713	446	552	476	786	681	549
Coefficient of Variation of Departures	.105	.070	.067	.049	.063	.105	.050
Coefficient of Variation of Employment	.070	.053	.086	.030	.035	.109	.035
Ratio of Coefficients of Variation–Employment/Departures	.67	.76	1.28	.62	.56	1.04	.70
Standard Error of Detrended-Deseasonalized Log Departures	.052	.065	.066	.052	.042	.058	.040
Standard Error of Detrended-Deseasonalized Log Employment	.035	.036	.041	.031	.036	.081	.031
Ratio of Standard Errors-Employment/Departures	.67	.55	.62	.60	.86	1.40	.78

Notes: Data pertain to domestic operations. Data from quarters with strike activity are removed. The following airlines had strikes during the sample period: American (1969:I), Continental (1976:IV), TWA (1973:IV), United (1975:IV) and Western (1969:III). All data are from CAB published and unpublished sources.

roughly ten quarters. As one might expect from Table 3, there is some heterogeneity across the airlines in the autoregressive representation of employment, particularly between United and the other six carriers. An F-test that the coefficients are the same across the seven airlines yields a probability value of just under .001.

Column 2 of Table 4 presents the coefficients of two lagged values of departures in explaining maintenance employment at the seven airlines. The coefficients are individually significant at conventional levels, although their joint contribution, summarized by the probability value in the last row (c), is only marginally significant. Columns 3 and 4 examine the role of lagged contract wages and lagged manufacturing wages in predicting employment. The evidence that contract wages influence employment is relatively weak, although the evidence for manufacturing wages is stronger. Finally, column 5 includes both wage measures together. In this equation, increases in contract wages have a significantly negative effect on employment, whereas increases in manufacturing wages lead to an eventual increase in employment.[13]

The results in column 5 are apparently robust to several alternative definitions of the opportunity wage for airline mechanics. In particular, if the manufacturing wage is replaced by the average wage of maintenance mechanics in manufacturing (interpolated from annual BLS *Area Wage Survey*

---

[13] A comparable regression of employment on contemporaneous departures, contract wages, and manufacturing wages yields a coefficient for contract wages of $-.034$ ($SE=.045$) and a coefficient for manufacturing wages of .061 ($SE=.138$).

TABLE 4—AUTOREGRESSIVE REPRESENTATIONS OF EMPLOYMENT[a]

	Dependent Variable: Logarithm of Maintenance Employment				
	(1)	(2)	(3)	(4)	(5)
Employment	1.16	1.09	1.09	1.09	1.08
$(t-1)$	(.06)	(.06)	(.06)	(.06)	(.06)
Employment	−.42	−.34	−.34	−.36	−.36
$(t-2)$	(.07)	(.07)	(.06)	(.06)	(.06)
Departures	–	.12	.13	.14	.13
$(t-1)$		(.04)	(.04)	(.04)	(.04)
Departures	–	−.10	−.09	−.11	−.09
$(t-2)$		(.04)	(.04)	(.04)	(.04)
Contract Wages	–	–	.02	–	−.01
$(t-1)$			(.04)		(.04)
Contract Wages	–	–	−.06	–	−.10
$(t-2)$			(.04)		(.04)
Manufacturing Wages	–	–	–	−.28	−.24
$(t-1)$				(.16)	(.16)
Manufacturing Wages	–	–	–	.46	.53
$(t-2)$				(.16)	(.16)
Probability Value of Exclusion Test for:					
(a) Departures	–	.05	.04	.04	.04
(b) Contract Wages	–	–	.27	–	.01
(c) Manufacturing Wages	–	–	–	.03	.02

Notes: All regressions include constants, trends, and quarterly dummy variables. Coefficients (except constants, trends, and seasonals) are restricted to be equal across airlines. Wage rates are deflated by the CPI. The probability values in the last row (a)–(c) refer to an F-test for the joint significance of two lagged values of the indicated variable.

[a] The standard errors are shown in parentheses.

data) or the wage rate for licensed mechanics at Boeing (available from the BLS Wage Chronology series), the coefficients are very similar in size and magnitude to those in Table 4, although less precisely estimated. I also constructed airline-specific measures of the opportunity wage based on earnings of maintenance mechanics and unemployment rates in cities where each airline has its major maintenance base.[14] These alternative wage measures give the same pattern of results as the manufacturing wage rate, although each can be rejected as a significant determinant of maintenance employment after controlling for the manufacturing wage rate. In this paper I therefore use the manufacturing wage as the alternative wage rate for airline mechanics.

Table 5 examines the contribution of several additional explanatory variables for maintenance employment. In each case, the employment equation contains two lagged values of the explanatory variable listed in the column heading, together with lagged values of employment, departures, contract wages, and manufacturing wages. The first four columns present employment equations that include alternative aggregate variables: consumer prices, real national income, an index of CAB-regulated passenger fares, a jet fuel price index, and an index of parts prices. None of these variables significantly improves the prediction of employment, maintaining lagged departures and wage rates. The next five columns report employment regressions that include additional airline-specific measures of output: available seat miles, revenue-passenger miles, domes-

---

[14] If the appropriate opportunity wage is the wage on an alternative job multiplied by the probability of being offered that job, then an observable proxy for the opportunity wage is $(1-U)a$, where U is the relevant unemployment rate and a is the wage rate of maintenance mechanics in the relevant city. This wage index was suggested by Brown and Ashenfelter.

TABLE 5—ADDITIONAL AUTOREGRESSIVE MODELS OF EMPLOYMENT

	Additional Explanatory Variable Included in Employment Regression								
	Consumer Prices (1)	Real GNP (2)	Output Price Index[a] (3)	Fuel Price Index[b] (4)	Parts Price Index[c] (5)	Available Seat Miles[d] (6)	Revenue Passenger Miles[e] (7)	Domestic Flight Hours[f] (8)	Total Flight Hours[g] (9)
Employment	1.03	1.06	1.09	1.07	1.09	1.00	1.07	1.04	1.08
$(t-1)$	(.06)	(.06)	(.06)	(.06)	(.07)	(.07)	(.06)	(.06)	(.06)
Employment	−.31	−.33	−.36	−.35	−.34	−.31	−.34	−.31	−.35
$(t-2)$	(.06)	(.06)	(.06)	(.06)	(.07)	(.06)	(.06)	(.07)	(.06)
Departures	.15	.13	.13	.11	.09	.25	.14	.21	.18
$(t-1)$	(.04)	(.04)	(.03)	(.04)	(.04)	(.06)	(.04)	(.08)	(.06)
Departures	−.11	−.09	−.07	−.08	−.07	−.05	−.11	−.11	−.09
$(t-2)$	(.04)	(.04)	(.03)	(.04)	(.04)	(.06)	(.04)	(.07)	(.06)
Contract Wages	.02	−.01	.01	−.03	.01	.03	−.02	.01	.00
$(t-1)$	(.04)	(.04)	(.03)	(.04)	(.04)	(.04)	(.04)	(.04)	(.03)
Contract Wages	−.10	−.10	−.11	−.07	−.11	−.10	−.08	−.10	−.08
$(t-2)$	(.03)	(.04)	(.03)	(.04)	(.04)	(.04)	(.04)	(.04)	(.04)
Manufacturing	−.53	−.26	−.34	−.14	−.39	−.41	−.26	−.22	−.24
Wages $(t-1)$	(.23)	(.17)	(.15)	(.20)	(.18)	(.18)	(.16)	(.19)	(.17)
Manufacturing	.74	.76	.48	.64	.73	.58	.59	.58	.53
Wages $(t-2)$	(.21)	(.20)	(.14)	(.20)	(.20)	(.18)	(.17)	(.19)	(.17)
Added Variable	−.33	−.08	.07	.01	.15	−.12	−.03	−.10	−.06
$(t-1)$	(.26)	(.11)	(.07)	(.03)	(.08)	(.04)	(.02)	(.07)	(.06)
Added Variable	.27	−.04	.01	.02	−.10	−.04	.02	.00	−.01
$(t-2)$	(.21)	(.10)	(.07)	(.03)	(.08)	(.04)	(.02)	(.07)	(.06)
Probability Value for Exclusion of Added Variable	.55	.34	.50	.34	.17	.01	.57	.22	.46

Notes: See notes to Table 4. The standard errors are shown in parentheses.

[a] Price index of domestic passenger fare rates set by CAB (constructed by the author).

[b] Price index for jet fuel (constructed by the author).

[c] Price index for commodity group 14: miscellaneous parts and machinery.

[d] Available seat miles on scheduled domestic passenger routes.

[e] Revenue passenger miles on scheduled domestic passenger routes.

[f] Aircraft revenue hours in domestic service, including nonscheduled, freight, and scheduled passenger service.

[g] Aircraft revenue hours in domestic and international service.

tic flight hours, and total flight hours (including domestic and international flight hours). Apart from available seat miles, none of these output measures is an important determinant of employment, once the level of departures is taken into account. Departures and available seat miles are highly colinear, however, and the sums of the coefficients on departures and seat miles are very nearly equal to the corresponding coefficients of departures when seat miles are excluded from the regression. Since the coefficients of the other variables are not much affected by the presence or absence of seat miles in the employment equation, I use departures as the sole measure of airline output in the remainder of the paper. Finally, while the results are not recorded in the table, I have also computed employment regressions that include airline-specific fleet composition variables.[15] Although these measures of firm-specific capital stock are marginally significant determinants of employment, their inclusion has no effect on the nature of the results in Tables 4 or 5.

The analysis of employment in Tables 4 and 5 is restricted to the number of mechan-

---

[15] Airline-specific fleet composition data are available on a annual basis from the Federal Aviation Administration *Statistical Handbook of Aviation*. For purposes of the employment regressions, I interpolated the number of aircraft of each type of airline by quarter, and grouped aircraft into five types.

TABLE 6—AUTOREGRESSIVE REPRESENTATIONS OF DEPARTURES

	Additional Explanatory Variable Included in Departures Regression:				
	None (1)	Employment (2)	Real GNP (3)	Manufacturing Wage (4)	Consumer Prices (5)
Departures	1.07	1.03	.91	1.00	1.00
$(t-1)$	(.07)	(.06)	(.06)	(.07)	(.07)
Departures	−.24	−.26	−.22	−.23	−.24
$(t-2)$	(.07)	(.06)	(.06)	(.07)	(.06)
Added Variable	–	.00	.24	.20	−.49
$(t-1)$		(.06)	(.16)	(.32)	(.33)
Added Variable	–	−.05	−.22	−.08	.49
$(t-2)$		(.06)	(.17)	(.32)	(.33)
Probability Value for Exclusion of Added Variable	–	.35	.38	.69	.38

Notes: See notes to Table 4. The standard errors are shown in parentheses.

ics on airline payrolls in each quarter. A more complete description of labor inputs, however, requires information on hours per employed worker. Although CAB records do not include a direct measure of hours per worker, a noisy indicator of hours is available from data on average payroll cost per worker. Specifically, the ratio of payroll cost per worker to the contract wage rate represents the sum of straight-time hours per worker, average overtime hours per worker (weighted by the overtime wage premium), and fringe benefit costs per worker (expressed as a fraction of the straight-time wage rate).[16] A regression of this hours index on contemporaneous employment and departures reveals no significant correlation with either variable.[17] This absence of correlation suggests that measured hours variation can be safely ignored in the study of employment and output.

The interpretation of employment equations that include lagged output variables, such as those in Tables 4 and 5, depends critically on the time-series representation of output. Table 6 presents several alternative representations of the level of departures activity at the seven airline firms. In each case the logarithm of departures is regressed on two lagged values of departures and two lagged values of the explanatory variable listed in the column heading. The coefficient estimates of a univariate forecasting equation (col. 1) show that departures have a monotonically declining moving-average representation, rather than the hump-shaped representation that characterizes employment. Column 2 presents the coefficients of lagged employment in a forecasting equation for departures, as well as the probability value of the associated exclusion test. The hypothesis that departures are exogenous to maintenance employment is not rejected at conventional significance levels. Columns 3–5 present the coefficients of three alternative aggregate variables in the departures equations: real output, manufacturing wages, and consumer prices. None of these is significantly related to departures activity, however, controlling for lagged departures.

---

[16] Let E represent payroll cost per worker, let w represent the union wage scale, let h_1 and h_2 represent average straight-time and overtime hours per worker, respectively, and let g represent average fringe benefit costs per worker, then

$$E = wh_1 + w(1+\omega)h_2 + g,$$

where ω represents the average overtime wage premium. The ratio of payroll cost to the contract wage rate is therefore

$$E/w = h_1 + (1+\omega)h_2 + g/w.$$

[17] The coefficient on the logarithm of employment is .06 ($SE = .06$) and the coefficient on the logarithm of departures is .02 ($SE = .05$).

To conclude this preliminary data analysis, the main conclusions may be summarized as follows:

1) The serial correlation properties of firm-specific employment and wage data for airline mechanics are very similar to the properties of aggregate data. Airline mechanics' real wage rates follow a first-order autoregressive process, while their employment levels follow a second-order process.

2) In the period under study, wage rates of airline mechanics were very similar across firms and uncorrelated with firm-specific employment levels. Real wage rates were significantly correlated with lagged manufacturing wage rates and lagged consumer prices.

3) Employment of airline mechanics is correlated with lagged values of contract wage rates and lagged values of wage rates outside the airline sector. The separate effects of these two wage rates are well-determined and in opposite directions. The employment effect of outside wage rates is apparently robust to alternative definitions of the outside wage, although average hourly earnings in manufacturing have the strongest correlation with employment of airline mechanics.

Building on these conclusions, the next section presents an intertemporal contracting model that provides a framework for testing between alternative models of the link between wage rates and employment.

II. Contractual Employment with Costly Adjustment

This section presents a simple extension of the static efficient contracting model to an intertemporal setting. Adjustment costs are introduced on the firm's side in order to generate an employment function with serial persistence and cumulative rather than instantaneous responses to changes in output or wages. The resulting function expresses optimal employment in terms of lagged values of employment, output, and wages, and provides a convenient framework for testing alternative models of employment determination. Both the labor demand model of employment determination, in which the employer takes contract wages as given, and the strong form efficiency model, in which the marginal product of labor is equated to the alternative wage rate, are obtained as special cases of the model. In contrast to tests of the efficient contract hypothesis based on a static employment function, the time-series correlations of employment and wages are modeled explicitly, and used to disentangle the forecasting role of alternative wages from any efficient contracting effects.

For simplicity, the dynamic relationship between output, wages, and employment is assumed to arise solely from the demand side of the contract. Workers are modeled as having separable preferences over time, with no interaction between lagged values of wages or employment and the ranking of current wage-employment pairs. While this assumption simplifies the analysis and interpretation of the dynamic employment function, it restricts the role of lagged wages and employment in generating current employment, and represents an obvious channel for further research.[18]

The first step in specifying the contractual employment function for airline mechanics is to specify the link between flight activity, maintenance activity, and employment. Airline mechanics service and inspect aircraft between departures ("line service") and also rebuild and overhaul aircraft components at major service intervals.[19] In either case, a variety of substitutes is available for in-house mechanics' services, including outside subcontractors and purchases of new parts and equipment. In addition, airlines can substitute mechanics' services over time by adding to or running down the stock of airworthy equipment. For simplicity, however, I assume that a given level of flight activity in the tth quarter, F_t, requires a proportional input of maintenance activities. I also assume that that maintenance is produced by a combination of in-house employment N_t and other inputs M_t, according to a Cobb-Douglas production function

---

[18] If workers bear mobility costs of moving between contract and alternative employment than presumably these costs should be internalized in an optimal employment contract. Alan Carruth and Andrew Oswald (1985) discuss the formulation of workers' objectives in an intertemporal contracting model.

[19] Each of the seven airlines in this study maintains a major maintenance depot where airframes and engines can be dismantled and rebuilt. In the industry as a whole, mechanics are more or less evenly split between line service and major maintenance activities.

$$F_t = AN_t^{\gamma_1}M_t^{\gamma_2},$$

where γ_1 and γ_2 are positive constants and A is a constant depending on aircraft type and route structure. The direct cost of maintaining a level of flight activity F_t with a labor force of N_t mechanics is therefore

$$(1) \qquad w_tN_t + q_tr(N_t, F_t),$$

where w_t is the contract wage for mechanics, q_t is the price of other maintenance inputs, and $r(N_t, F_t) = A^{-1/\gamma_2}N_t^{-\gamma_1/\gamma_2}F_t^{1/\gamma_2}$ is the input requirement function for nonlabor inputs, given labor inputs N_t and output F_t.

In addition to these direct costs, I assume that the firm bears an adjustment cost $J(N_t, N_{t-1})$ in period t that depends on the level of employment in t and $t-1$. This adjustment cost captures both hiring and firing costs, and the cost of rearranging flight schedules as the number of employees available for line service at each airport is adjusted over time.

The final ingredient of the employment contract is the specification of workers' preferences. I assume that preferences in each period are represented by a function of contract wages, employment, outside or alternative wages a_t, and a random preference shock ν_t: $U(N_t, w_t, a_t, \nu_t)$. Two functional forms are considered for U. In the first case, assume that N_0 workers are attached to the firm, and that workers are allocated randomly in each period between contract employment and alternative employment with probabilities N_t/N_0 and $1 - N_t/N_0$, respectively. If $v(x, \nu_t)$ is a von Neumann-Morgenstern utility function defined on the level of earnings x, then preferences of a representative worker are summarized by

$$(2) \quad U(N_t, w_t, a_t, \nu_t) = \frac{N_t}{N_0}v(w_t, \nu_t) + \left(1 - \frac{N_t}{N_0}\right)v(a, \nu_t).$$

In the second case, following John Pencavel (1984), assume that workers' preferences can be summarized by a Cobb-Douglas function of employment and the gap between contract and alternative wages:

$$(3) \qquad U(N_t, w_t, a_t, v_t) = k(\nu_t)N_t^{\theta_1}(w_t - a_t)^{\theta_2}.$$

Both of these specifications contain as a special case the "excess earnings" objective $N_t(w_t - a_t)$ associated with an income-maximizing union.[20]

Under the assumption that current and future flight activity are exogenous to maintenance employment, an optimal contract minimizes the expected present value of employer's costs, subject to an expected utility requirement for workers. Assuming that employers and workers have a constant discount rate $\beta (\beta < 1)$, the level of employment and wages in period t solves

$$(4) \quad \min_{w_t, N_t} E \sum_{j=0}^{\infty} \beta^j \big[w_{t+j}N_{t+j} + q_{t+j}r(N_{t+j}, F_{t+j}) + J(N_{t+j}, N_{t+j-1})\big],$$

subject to

$$E \sum_{j=0}^{\infty} \beta^j U(N_{t+j}, w_{t+j}, a_{t+j}, \nu_{t+j}) \geq U_0.$$

---

[20] For the expected utility specification, set $v(x_t, \nu_t) = x_t$. For the Cobb-Douglas specification, set $k(\nu_t) = 1$ and $\theta_1 = \theta_2 = 1$. Income maximization is an appropriate objective for workers as a group if there are no constraints on the internal distribution of earnings or employment opportunities among the group.

In these expressions expectations (denoted E) are taken over the joint distribution of the entire sequence of future flight activity, input prices, alternative wages, and preference shocks. The solution to this constrained-optimization problem can be obtained as the solution to the Lagrangian expression

$$(5)\quad \min_{w_t, N_t} E \sum_{j=0}^{\infty} \beta^j \big[w_{t+j} N_{t+j} + q_{t+j} r(N_{t+j}, F_{t+j}) + J(N_{t+j}, N_{t+j-1}) - \mu U(N_{t+j}, w_{t+j}, a_{t+j}, \nu_{t+j}) \big],$$

for some positive constant μ.

In contrast to the efficient contracting model, which views employment and wages as jointly determined, the traditional labor demand model treats employment as determined unilaterally by the firm, taking contract wages as given. In this case, observed employment solves the employer's cost-minimization problem

$$(6)\quad \min_{N_t} E \sum_{j=0}^{\infty} \beta^j \big[w_{t+j} N_{t+j} + q_{t+j} r(N_{t+j}, F_{t+j}) + J(N_{t+j}, N_{t+j-1}) \big],$$

directly, subject to the forecasting equations for flight activity, input prices, and contract wages. A comparison of equations (5) and (6) reveals two important differences between the efficient contract and labor demand models. First, the efficient contract model treats wages and employment as jointly endogenous. Second, while alternative wages enter the contracting model directly through workers' preferences, in the labor demand model the only role of alternative wages is in the forecasting equation for future contract wages.

Before deriving the employment functions associated with the contracting and labor demand models, it is useful to characterize the wage function implied by the efficient contracting model. In particular, it is interesting to ask if a contracting model which considers wages and employment as jointly determined can ever lead to the prediction that wage outcomes are independent of previous employment levels. The empirical analysis in Section I suggests that this is an important characteristic of contract wage rates for airline mechanics.

The first-order condition for wages in period $t+j$ for the contracting model (5) is

$$(7)\quad N_{t+j} - \mu U_w(N_{t+j}, w_{t+j}, a_{t+j}, \nu_{t+j}) = 0.$$

From this equation it is evident that the choice of contract wages in $t+j$ is independent of employment if workers' preferences are *linear* in employment: say,

$$U(N, w, a, \nu) = N f(w, a, \nu).$$

In that case, the first-order condition (7) has the simple form

$$\mu f_w(w_{t+j}, a_{t+j}, \nu_{t+j}) = 1,$$

with the implication that contract wages are determined in each period independent of employment or wage choices in any other period. Since the evidence in Section I suggests that contract wages are unrelated to past employment, the assumption that workers' preferences are linear in employment is plausible as well as convenient, and will be adopted here. The expected-utility preference specification (3) imposes linearity a priori.[21] For the Cobb-Douglas specification (4), linearity implies a within-period objective of the form $N_t(w_t - a_t)^{\theta}$, where $\theta = \theta_2 > 0$ may be either greater than or less than unity.

---

[21] The expected utility preference specification is linear in employment under the assumption that the probability of employment is proportional to the actual level of employment.

Specializing the first-order condition (7) to these two preference specifications leads to

$$\mu v_w(w_{t+j}, \nu_{t+j}) = 1, \tag{8a}$$

in case of expected utility preferences, and

$$k(\nu_{t+j})\mu\theta(w_{t+j} - a_{t+j})^{\theta-1} = 1, \tag{8b}$$

in case of Cobb-Douglas preferences. The expected utility specification implies that real contract wages are constant over time, apart from changes in the preference shock ν or random measurement errors. The Cobb-Douglas specification, on the other hand, suggests that contract wages maintain a constant (absolute) differential over alternative wages. For example, if $k(\nu_t) = \nu_t^{1-\theta}$ and ν_t is first-order autoregressive, then equation (8b) implies

$$w_t = \delta_1 w_{t-1} + a_t - \delta_1 a_{t-1} + \nu_t',$$

where δ_1 is the first-order autocorrelation coefficient of ν_t and $\nu_t' = \nu_t - \delta_1 \nu_{t-1}$ is serially uncorrelated. Comparison of this equation with the fitted regressions in Section I suggests that the first-order condition (8b) may provide a useful model of the contract wage determination process for airline mechanics.

The assumption that workers' preferences over employment and wage outcomes are linear in employment simplifies the analysis of contractual employment setting. Let w_{t+j}^* represent the solution to equation (8a) or (8b). Since contract wages are unrelated to past employment decisions, the optimal level of contract employment in t solves

$$\min_{N_t} E \sum_{j=0}^{\infty} \beta^j \Big[w_{t+j}^* N_{t+j} + q_{t+j} r(N_{t+j}, F_{t+j}) + J(N_{t+j}, N_{t+j-1}) - \mu U(N_{t+j}, w_{t+j}^*, a_{t+j}, \nu_{t+j}) \Big] \tag{9}$$

subject to the forecasting equations for flight activity, alternative wages, preference shocks, and optimized contract wages.[22] If employment fails to Granger-cause wages, the contract employment function can be obtained by taking the contract wage as exogenous. In this case, the labor demand model of employment determination (6) is a special case of the contracting model (9) with $\mu = 0$.

In order to derive the contract employment function when there are costs of changing employment from period to period, it is convenient to proceed in two steps. The first is to derive the optimal employment level in the absence of adjustment costs. The second is to derive the actual employment decision by comparing the costs of changing employment over time with the cost of sub- or superoptimal employment in each period.[23] For simplicity, these two cost components are expressed as quadratic functions of the logarithm of employment, yielding an employment function that is linear in the logarithms of employment, output, and wages.

The optimal employment choice in the absence of adjustment cost can be obtained from the first-order condition for the contracting problem (9), setting the adjustment cost component to zero:

$$w_{t+j} + q_{t+j} r_N(N_{t+j}, F_{t+j}) - \mu U_N(N_{t+j}, w_{t+j}, a_{t+j}, \nu_{t+j}) = 0. \tag{10}$$

For the expected utility specification of workers' preferences, this first-order condition implies

$$-q_{t+j} r_N(N_{t+j}, F_{t+j}) = w_{t+j} - \mu\big(v(w_{t+j}) - v(a_{t+j})\big),$$

where for simplicity I have suppressed the dependence of v on the preference shock ν_t. The term on the left-hand side of this equation represents the marginal value product of labor, measured by the savings in nonlabor input costs as employment is increased by one unit. The marginal value product of labor is equated to the current contract wage, minus a premium that depends on the gap between contract and alternative wages. Using the properties of the optimal contract wage (equation (8a)), the appropriate shadow value of mechanics' labor is

---

[22] Since w_t is a function of a_t and ν_t, say $w_t = g(a_t, \nu_t)$, rational forecasts for w_{t+j}^* satisfy $Ew_{t+j}^* = Eg(a_{t+j}, \nu_{t+j})$.

[23] John Kennan (1979) provides a useful discussion of this two-step procedure.

$$S_{t+j} \equiv w_{t+j} - \mu\big(v(w_{t+j}) - v(a_{t+j})\big)$$
$$= w_{t+j}\left(1 - \frac{v(w_{t+j}) - v(a_{t+j})}{w_{t+j} v_w(w_{t+j})}\right).$$

If the expected utility function v is linear in earnings, then the expression for the shadow value of labor reduces to a_{t+j}, the alternative wage rate. More generally, using a second-order expansion for $v(a_{t+j})$ around w_{t+j}, the shadow value of labor in the absence of adjustment costs is approximately

$$\text{(11a)} \quad S_{t+j} \cong a_{t+j}(1 + \delta\pi) - \delta\pi w_{t+j},$$

where δ is the coefficient of relative risk aversion of the utility function $v(\delta \geq 0)$ and π represents the average markup of the optimal contract wage over the alternative wage.[24] If workers' preferences have the expected utility form, the shadow value of labor in an optimal contract is decreasing in the contract wage. As noted by Ian McDonald and Robert Solow (1981), this specification of workers' preferences implies that efficient combinations of employment and wages are positively correlated across otherwise identical contracts.

For the case of Cobb-Douglas preferences, the first-order condition for employment in the absence of adjustment costs is

$$-q_{t+j} r_N(N_{t+j}, F_{t+j})$$
$$= w_{t+j} - \mu k(\nu_{t+j})(w_{t+j} - a_{t+j})^{\theta}.$$

Again, the shadow value of labor in the optimal contract is lower than the contract wage. Using the first-order condition (8b) for the optimal contract wage in period $t + j$, the shadow value of labor is

$$\text{(11b)} \quad s_{t+j} = a_{t+j}(1/\theta) + w_{t+j}(1 - (1/\theta)).$$

For values of θ less than unity, this expression is identical to (11a). For values of θ in excess of unity, however, the Cobb-Douglas specification implies that S_t is increasing in both the contract and the alternative wage.

In view of the similarity of (11a) and (11b), it is straightforward to derive the optimal level of employment in period t for either specification of workers' preferences. Assuming that arithmetic and geometric averages of contract and alternative wages are equal, the logarithm of the appropriate shadow value of labor is

$$\text{(12)} \quad \log S_t = \alpha \log w_t + (1 - \alpha)\log a_t,$$

where $\alpha = 1 - 1/\theta$ in the Cobb-Douglas specification of workers' preferences, and $\alpha = -\pi\delta$ in the expected utility specification. In a strong form efficient contract where workers' objectives are summarized by the value of "excess earnings," $N_t(w_t - a_t)$, $\alpha = 0$. In the labor demand model, on the other hand, the relevant opportunity cost of labor is the contract wage and $\alpha = 1$. Substi-

---

[24] Dropping time subscripts, expand $v(a) = v(w) + (a - w)v'(w) + \frac{1}{2}(a - w)^2 v''(w)$, and write

$$S = w\left(1 - \frac{v(w) - v(a)}{wv'(w)}\right),$$
$$= w\left(1 - \frac{w - a}{w} - \frac{\delta}{2}\left(\frac{w - a}{w}\right)^2\right)$$

where $\delta = -wv''(w)/v'(w)$. Next, linearize $((w - a)/w)^2$ around π, where π is the average markup of w over a:

$$((w - a)/w)^2 \cong 2\pi((w - a)/w) - \pi^2.$$

Substituting this expression into the expression for S and assuming that π^2 is neglible yields $S = a(1 + \delta\pi) - w\delta\pi$.

tuting equation (12) into the first-order condition for employment in the absence of adjustment cost and taking logarithms yields the optimal employment level

$$(13)\quad \log N_t^* = \text{constant} + b_1 \log F_t + b_2 \log w_t + b_3 \log a_t + b_4 \log q_t,$$

where $b_1 = 1/(\gamma_1+\gamma_2)$, $b_2 = -\alpha\gamma_2/(\gamma_1+\gamma_2)$, $b_3 = -(1-\alpha)\gamma_2/(\gamma_1+\gamma_2)$, and $b_4 = \gamma_2/(\gamma_1+\gamma_2)$. In effect, N_t^* is the level of employment observed on the firm's labor demand curve when output is F_t and the wage rate is a weighted average of the alternative wage a_t and the contractual wage w_t.[25]

The cost of maintaining a level of flight activity F_t with a labor force $N_t \neq N_t^*$ can be obtained from a second-order expansion of the appropriate cost function. Let

$$c(N_t) = w_t N_t + q_t r(N_t, F_t) - \mu U(N_t, w_t, a_t)$$

denote the cost (net of the contribution to workers' utility) of maintenance activities in quarter t, excluding adjustment costs. Using the fact that $c'(N_t^*) = 0$,

$$\begin{aligned} c(N_t) &\cong c(N_t^*) + (N_t - N_t^*)c'(N_t^*) \\ &\quad + \tfrac{1}{2}(N_t - N_t^*)^2 c''(N_t^*) \\ &\cong c(N_t^*) + \tfrac{1}{2}c_{1t}(\log N_t - \log N_t^*), \end{aligned}$$

where c_{1t} represents a second-order expansion coefficient when deviations of N_t from N_t^* are taken in proportionate terms.[26] I assume that this approximation continues to hold when c_{1t} is replaced by its sample average value c_1.

The optimal employment choice in period t can be obtained by combining the preceding expression with the adjustment cost terms $J(N_{t+j}, N_{t+j-1})$. For convenience I assume that the costs of changing the labor force are related to the proportional change in employment by

$$J(N_{t+j}, N_{t+j-1}) = \tfrac{1}{2}c_2(\log N_{t+j} - \log N_{t+j-1})^2.$$

The assumption of equal adjustment costs for equiproportional increases and decreases in employment is particularly restrictive, but is required for empirical tractibility.

With this setup, it is straightforward to derive the dynamic employment equation for airline mechanics. The choice of current employment minimizes the following quadratic expression:

$$(14)\quad E\sum_{j=0}^{\infty}\beta^j\Big[c(N_{t+j}^*) + \frac{1}{2}c_1(\log N_{t+j} - \log N_{t+j}^*)^2 + \frac{1}{2}c_2(\log N_{t+j} - \log N_{t+j-1})^2\Big].$$

The solution to this class of problems is well-known and can be summarized by the partial-adjustment equation

$$(15)\quad \log N_t = \lambda \log N_{t-1} + (1-\lambda)(1-\lambda\beta)\sum_{j=0}^{\infty}(\lambda\beta)^j E_t \log N_{t+j}^*,$$

where λ is a root to the quadratic equation

$$\lambda^2 - \left(\frac{1+\beta+c_1/c_2}{\beta}\right)\lambda + \frac{1}{\beta} = 0$$

lying between 0 and 1. According to equation (15), observed employment represents a weighted average of last period's employment and the discounted average of expected future values of N_t^*. The adjustment parameter λ reflects the relative size of the coefficients c_1 and c_2. The larger is c_2, the more costly is labor force adjustment, and the larger is λ.[27]

---

[25] Equation (13) can also be written in terms of alternative wages and the preference shock variable v_t using the first-order condition for contract wages. For purposes of forming forecasts of future values of N_t^*, however, it is convenient to express the employment equation in terms of w_t, since the forecasting equation for w_t is directly observable.

[26] In the Cobb-Douglas case, c_{1t} is proportional to the value of nonlabor inputs at the optimal level of labor inputs.

The solution for N_t can be obtained by substituting from equation (13) for $\log N_t^*$ into (15). The resulting expression translates $\log N_t$ into a function of $\log N_{t-1}$, and discounted averages of expected future values of $\log F_t$, $\log w_t$, $\log a_t$, and $\log q_t$. These expressions can in turn be written as functions of current and past values of wages and departures, and current and past values of all variables useful in predicting wages or departures.

The presence of unobserved error components in the static employment function (13) introduces an additional consideration into the formulation of the dynamic employment function. Suppose that equation (13) contains a stochastic productivity effect ε_t. Then, excluding constants, the dynamic employment function is

$$(16)\quad \log N_t = \lambda \log N_{t-1} + (1-\lambda)(1-\lambda\beta) \times \sum_{j=0}^{\infty} (\lambda\beta)^j E_t\big[b_1 \log F_{t+j} + b_1 \log w_{t+j} + b_3 \log a_{t+j} + b_4 \log q_{t+j} + \varepsilon_{t+j}\big].$$

An empirically useful hypothesis is that the error component ε_t is first-order autoregressive.[28] In particular, let

$$\varepsilon_t = \rho\varepsilon_{t-1} + \xi_t,$$

where ξ_t is serially uncorrelated, and strictly exogenous to flight activity, employment or wages. Then

$$(1-\lambda\beta)\sum_{j=0}^{\infty}(\lambda\beta)^j E_t \varepsilon_{t+j} = \frac{1-\lambda\beta}{1-\lambda\beta\rho}\varepsilon_t,$$

which introduces a first-order autoregressive error into the dynamic employment function (16), and generates a second-order autoregressive representation of employment.

The terms on the right-hand side of equation (16) depend on the definitions and forecasting equations for alternative wages, flight activity, nonlabor input prices, and contract wages. On the basis of the evidence reported in Section I, I assume that the alternative wage for airline mechanics is represented by the average hourly wage rate in manufacturing (apart from trend and seasonal factors). As a measure of flight activity, I use the level of domestic departures. Finally, for lack of suitable data, I assume that nonlabor input prices (q_t) are captured by trend and seasonal factors. This assumption is especially problematic if the main substitute for in-house employment is subcontracted maintenance, and if the price of subcontracted maintenance is correlated with mechanics' wage rates.

Building on the results in Section I, I adopt the following second-order autoregressive forecasting system for detrended and deseasonalized departures, manufacturing wages, contract wages, and consumer prices:

$$(17a)\quad \log F_t = \phi_1 \log F_{t-1} + \phi_2 \log F_{t-2} + u_{1t},$$

$$(17b)\quad \log a_t = \alpha_1 \log a_{t-1} + \alpha_2 \log a_{t-2} + \alpha_3 \log p_{t-1} + \alpha_4 \log p_{t-2} + u_{2t},$$

$$(17c)\quad \log w_t = \delta_1 \log w_{t-1} + \delta_2 \log w_{t-2} + \psi_1 \log a_{t-1} + \psi_2 \log a_{t-2} + \psi_3 \log p_{t-1} + \psi_4 \log p_{t-2} + u_{3t},$$

$$(17d)\quad \log p_t = \pi_1 \log p_{t-1} + \pi_2 \log p_{t-2} + \pi_3 \log a_{t-1} + \pi_4 \log a_{t-2} + u_{4t},$$

---

[27] The adjustment parameter λ is related to the ratio of c_1 to c_2 by $c_1/c_2 = (1-\lambda)(1-\lambda\beta)\lambda^{-1}$.

[28] Sargent suggests this hypothesis as a means of generating a second-order autoregressive model for employment from a cost-of-adjustment model.

where the vector of residuals $(u_{1t}, u_{2t}, u_{3t}, u_{4t})$ is assumed to be serially uncorrelated. In this forecasting system, aggregate prices and manufacturing wages depend on their own lagged values and lagged values of each other, while departures are forecast by a univariate second-order autoregression. Contract wages depend on their own lagged values as well as lagged values of manufacturing wages and consumer prices. The forecasting equation for contract wages can be interpreted as a log-linear approximation to the first-order condition (7) for optimal contract wages, although I do not restrict the coefficients of the equation in any way. The evidence that alternative wages help forecast contract wages is more consistent with the Cobb-Douglas preference specification than the expected utility specification, although neither model provides a ready interpretation of the role of prices in forecasting contract wage rates.[29]

The system of second-order forecasting equations (17) generate the following expressions:

$$(18a)\quad (1-\lambda\beta)\sum_{j=0}^{\infty}(\lambda\beta)^j E_t \log F_{t+j} = B_{11}\log F_t + B_{12}\log F_{t-1},$$

$$(18b)\quad (1-\lambda\beta)\sum_{j=0}^{\infty}(\lambda\beta)^j E_t \log a_{t+j} = B_{21}\log a_t + B_{22}\log a_{t-1} + B_{23}\log p_t + B_{24}\log p_{t-1},$$

$$(18c)\quad (1-\lambda\beta)\sum_{j=0}^{\infty}(\lambda\beta)^j E_t \log w_{t+j} = B_{31}\log w_t + B_{32}\log w_{t-1} + B_{33}\log a_t + B_{34}\log a_{t-1} + B_{35}\log p_t + B_{36}\log p_{t-1},$$

where the B_{ij} are known functions of the coefficients in equations (17a)–(17d). Substituting these equations into (16), performing a transformation to eliminate serial correlation in the productivity shock ε_t, and using (17a)–(17d) to substitute for current values of departures, prices and wages in terms of lagged values and innovations in these variables yields the reduced-form employment equation implied by the model. This reduced form contains two lagged values of employment, and each of the exogenous variables (including contract wages), as well as a residual that is a combination of the unanticipated productivity shock ξ_t and the current forecast errors u_{1t}, u_{2t}, u_{3t}, and u_{4t}. The coefficients of the reduced-form equation depend on the coefficients of equations (16) and (18), the adjustment parameter λ, and the serial correlation coefficient ρ of the unobserved productivity shock.[30] The dynamic properties of employment, however, depend solely on λ and ρ. The first-order autoregressive coefficient of employment is the sum of λ and ρ, while the second-order autoregressive coefficient is the negative of their product.

To illustrate the implications of the model for the reduced-form employment equation, it is useful to consider two polar models of employment determination: the labor demand model in which the relevant opportunity cost of contract labor is the contract wage; and the strong form efficient contract model in which the relevant opportunity cost is the alternative wage rate. In the labor demand model, $b_3 = 0$ in equation (16) and the alternative wage effects employment only through the forecasting equation for contract wages. If future contract wages depend positively on manufacturing wages, for example, then employment should depend negatively on manufacturing wages, at least in the long run. The evidence in Tables 2 and 4, however, shows that manufacturing wages have a positive long-run impact on contract wages, and a *positive* long-run effect on employment. In the strong form efficient contract, on the other hand, N_t^* is independent of contract wages and depends only on the level of alternative wages. This model is also rejected by the unrestricted employment regressions in Table 4, which show a negative impact of contract wages on employment, holding constant manufacturing wages. More formal tests between these two polar models, and tests of the overidentifying restrictions implied by the general contracting model, are presented in the next section.

---

[29] The theoretical model assumes that contract wage rates are adjusted every period, whereas mechanics wage rates are set in two- or three-year nominal contracts. This suggests that there may be a cost of adjusting contract wage rates that is missing from the model.

[30] A detailed derivation of the reduced-form employment equation is presented in Appendix I.

III. Empirical Analysis of the Dynamic Contracting Model

The model of employment determination developed in Section II consists of the prediction equations for contract wages, alternative wages and departures (equations (17a)–(17d)), together with the reduced-form employment equation implied by (16). This section presents estimation results based on fitting this five-equation system to aggregate quarterly data on manufacturing wages and consumer prices as well as firm-specific data on contract wages, employment, and departures for the seven airline firms. For simplicity, the firm-specific data are deseasonalized and detrended prior to estimation. This permits unrestricted airline-specific constants, trends, and seasonals to be fitted outside of the main estimation step, at the cost of some potential bias in the estimated standard errors.

In addition, rather than estimate equations for the airline-specific data and the aggregate data simultaneously, I have estimated the manufacturing wage and consumer price equations separately over a longer sample period (1964:III–1978:IV), and then used the estimated parameters as known constants in the calculation of the restricted employment equation. The model is therefore treated as a three-equation system for employment, wages, and departures, with *known* forecasting equations for aggregate wages and prices.

The employment, wage, and departures equations for each of the seven airline firms (21 equations in all) are fitted to detrended and deseasonalized data by a two-step nonlinear generalized least squares procedure.[31] The second-stage estimates minimize the weighted residuals of the twenty-one-equation system, using as weights the inverse covariance matrix formed by the unrestricted least-squares residuals. Following A. Ronald Gallant and Dale Jorgenson (1979), a goodness-of-fit test is constructed by comparing the weighted sum of squares of the restricted model to the weighted sum of squares of the unrestricted model.

Unrestricted vector-autoregressive representations of airline-specific departures, wages, and employment are presented in the first two columns of Table 7.[32] Column 1 contains estimates of the three-equation system when lagged prices are excluded from the wage and employment equations, while column 2 contains estimates of the system when lagged prices are included in these two equations. The coefficient estimates are very similar to the corresponding estimates in Tables 2, 4, and 6 obtained by fitting the system equation-by-equation to unadjusted data.

Columns 3 and 4 contain the restricted reduced-form parameter estimates associated with the labor demand version of equation (16). The corresponding structural parameter estimates are presented in columns 1 and 2 of Table 8. In this version of the contracting model, the relevant opportunity cost of labor is the contract wage rate, and manufacturing wages enter the employment equation for airline mechanics only insofar as they help

---

[31] As an alternative to this estimation procedure, the model can also be estimated by applying instrumental variables techniques to the first-order condition for contract employment. The instrumental variables procedure has the advantage that closed-form solutions for the optimal employment choice are not required. The explicit solution procedure adopted here, on the other hand, has the advantage of offering a direct interpretation of the reduced-form system for departures, wages, and employment.

[32] These unrestricted representations incorporate a variety of exclusion restrictions (for example, lagged employment is not included in the departures equation). Conditional on these exclusion restrictions, however, the estimates are unrestricted.

TABLE 7—REDUCED-FORM PARAMETER ESTIMATES FOR DEPARTURES, WAGES, EMPLOYMENT: SEVEN DOMESTIC AIRLINES 1969:III–1976:IV[a]

			Restricted Reduced Forms[b]				
	Unrestricted Reduced Forms		Labor Demand Model		Strong Form Efficient Contract Model		General Model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Departures Equation:							
Departures	1.08	1.06	1.09	1.10	1.09	1.10	1.13
$(t-1)$	(.16)	(.16)	(.15)	(.14)	(.15)	(.13)	.12
Departures	−.27	−.26	−.30	−.28	−.30	−.29	−.33
$(t-2)$	(.13)	(.13)	(.11)	(.14)	(.11)	(.09)	(.09)
Wage Equation:							
Contract Wages	.62	.62	.62	.64	.62	.64	.68
$(t-1)$	(.17)	(.16)	(.17)	(.13)	(.17)	(.12)	(.15)
Contract Wages	−.08	−.12	−.07	−.14	−.08	−.14	−.19
$(t-2)$	(.16)	(.15)	(.16)	(.11)	(.16)	(.10)	(.12)
Manufacturing Wages	1.43	.70	1.39	.59	1.37	.58	.65
$(t-1)$	(.77)	(.99)	(.76)	(.63)	(.73)	(.59)	(.71)
Manufacturing Wages	−.59	−.85	−.62	−.80	−.57	−.85	−.50
$(t-2)$	(.77)	(.96)	(.70)	(.61)	(.66)	(.55)	(.69)
Consumer Prices	–	−1.21	–	−1.22	–	−1.25	−1.01
$(t-1)$		(1.12)		(1.68)		(.82)	(.90)
Consumer Prices	–	.62	–	.63	–	.63	.56
$(t-2)$		(.96)		(.60)		(.71)	(.78)
Employment Equation:							
Employment	1.01	1.01	1.04	1.05	1.03	1.05	1.07
$(t-1)$	(.13)	(.14)					
Employment	−.29	−.30	−.26	−.27	−.26	−.27	−.28
$(t-2)$	(.11)	(.12)					
Departures	.15	.14	.15	.13	.14	.13	.15
$(t-1)$	(.07)	(.07)					
Departures	−.10	−.10	−.09	−.08	−.09	−.08	−.09
$(t-2)$	(.06)	(.07)					
Contract Wages	.05	.05	−.01	.00	0.0	0.0	.00
$(t-1)$	(.07)	(.08)					
Contract Wages	−.15	−.14	.00	−.01	0.0	0.0	−.02
$(t-2)$	(.07)	(.07)					
Manufacturing Wages	−.35	−.30	.02	.00	.06	.00	.01
$(t-1)$	(.27)	(.40)					
Manufacturing Wages	.67	.70	−.01	−.01	−.02	.00	−.06
$(t-2)$	(.29)	(.36)					
Consumer Prices	–	.07	–	−.02	–	.01	.01
$(t-1)$		(.50)					
Consumer Prices	–	−.03	–	.01	–	.01	−.07
$(t-2)$		(.37)					
Log Likelihood	−209.85	−194.03	−225.09	−208.08	−225.32	−206.09	−204.26

[a] Estimated on detrended and deseasonalized data. Observations from strike and immediate poststrike periods are deleted. Estimates are from the second stage of a two-step generalized least squares procedure. The standard errors are shown in parentheses.

[b] Restricted reduced forms are conditional on parameter estimates for consumer price and manufacturing wage equations. The estimates in cols. 3 and 4 correspond to the structural estimates in cols. 1 and 2 of Table 8. The estimates in cols. 5 and 6 correspond to the structural estimates in cols. 3 and 4 of Table 8. The estimates in col. 7 correspond to the structural estimates in col. 5 of Table 8.

TABLE 8—PARAMETER ESTIMATES FOR PARTIAL-ADJUSTMENT EMPLOYMENT EQUATION: SEVEN DOMESTIC AIRLINES 1969:III–1976:IV[a]

	Labor Demand Model		Strong Form Efficient Contract Model		General Model
	Prices Excluded (1)	Prices Included (2)	Prices Excluded (3)	Prices Included (4)	Prices Included (5)
Partial-Adjustment	.44	.44	.44	.44	.49
Parameter (λ)	(.22)	(.21)	(.22)	(.21)	(.24)
Serial Correlation	.60	.61	.60	.60	.57
Parameter (ρ)	(.23)	(.22)	(.23)	(.22)	(.26)
Output Elasticity	.77	.66	.70	.67	.76
(b_1)	(.41)	(.30)	(.37)	(.29)	(.30)
Contract Wage	.03	.03	0.0	0.0	.38
Elasticity (b_2)	(.05)	(.03)			(.91)
Alternative Wage	0.0	0.0	.29	−.05	−1.02
Elasticity (b_3)			(1.34)	(.95)	(2.44)
Log Likelihood	−225.09	−208.08	−225.32	−206.09	−204.26
Goodness-of-Fit					
against Unrestricted	30.48	28.10	30.94	24.12	20.46
Model[b]	[.000]	[.000]	[.000]	[.000]	[.001]

[a] See notes to Table 7. The standard errors are shown in parentheses. Estimates in cols. 1 and 3 are based on a forecasting equation for contract wages that excludes consumer prices. Estimates in cols. 2, 4, and 5 are based on a forecasting equation for contract wages that includes manufacturing wages and consumer prices. Estimates are conditional on parameter estimates for consumer price and manufacturing wage equations.

[b] Probability value is shown in brackets. The models in cols. 1 and 3 have 4 degrees of freedom. The models in cols. 2 and 4 have 6 degrees of freedom. The model in col. 5 has 5 degrees of freedom.

to predict future contract wages. The model is therefore an application of Thomas Sargent's dynamic employment demand model to firm-specific data, with output taken as exogenous. The two alternative specifications of the labor demand model differ by whether or not lagged prices are used to forecast future contract wages and future alternative wages.[33]

The estimated wage elasticities (parameter b_2) for the labor demand model are small and positive, and insignificantly different from zero. A comparison of the restricted reduced forms in columns 3 and 4 with the unrestricted reduced forms in columns 1 and 2 suggests several difficulties with the restricted fit. First, the labor demand model cannot explain the opposite signs of contract and manufacturing wages in the reduced-form employment equation, given that future contract wages are positively correlated with manufacturing wages. Second, in the specification of the model that includes prices in the forecasting equation for contract wages, the relatively small effects of prices on employment are difficult to reconcile with

---

[33] To estimate the reduced-form employment equation, I set the quarterly discount rate β to .99. When prices are included in the forecasting equations for wages, I use the following forecasting equations for aggregate manufacturing wages (a_t) and consumer prices (p_t):

$$\log a_t = .70\log a_{t-1} + .02\log a_{t-2} - .52\log p_{t-1} + .47\log p_{t-2}$$

$$\log p_t = 1.79\log p_{t-1} - .78\log p_{t-2} - .11\log a_{t-1} + .02\log a_{t-2}.$$

When prices are excluded from the wage forecasting equation, I use the following forecasting equation for manufacturing wages:

$$\log a_t = 1.09\log a_{t-1} - .26\log a_{t-2}.$$

the relatively large effects of prices on expected future contract wages. The goodness-of-fit statistics in the last row of Table 8 suggest that the labor demand interpretation of the employment-wage-output system is strongly rejected by the data.

In contrast to the poor performance of the model in summarizing the effects of wages on employment, the linkage between departures and employment is more successfully explained. The estimated (long-run) output elasticity of employment is between .60 and .80 and is not significantly different from unity. The AR(2) structure of employment is also apparently well-captured by the combination of adjustment costs and first-order serial correlation in demand: a comparison of the restricted and unrestricted reduced-form autoregressive parameters reveals only small differences between them.

Parameter estimates for the strong form efficient contract model, which takes the alternative wage rate as the opportunity cost of labor and excludes contract wages from the employment equation altogether, are presented in columns 3 and 4 of Table 8. The associated reduced forms are contained in columns 5 and 6 of Table 7. Again, a comparison of the unrestricted and restricted reduced forms shows that the model has trouble explaining the effects of manufacturing wages on employment. When prices are not used to forecast future manufacturing wage rates, the implied estimate of the elasticity of demand is .29. When prices are included in the wage forecasting equation, on the other hand, the estimated elasticity of demand is $-.05$. As is the case for the labor demand model of employment determination, the dynamic link between wages and employment is not well explained by the strong form efficient contract model, although the departures-employment relationship is reasonably well explained by either model.

Finally, estimates of a general contracting model that permits the shadow value of labor to depend on a weighted average of contract and alternative wages are presented in the last columns of Table 7 and Table 8. In this general model, prices are included in the forecasting equations for contract wages and manufacturing wages. A version of the model that excluded prices from the wage equations proved to be unidentified.[34] Although the general contracting model fits better than either polar model, the estimated wage elasticities of employment are poorly determined and not significantly different from zero. Again, the implied reduced-form coefficients of contract and manufacturing wages in the employment equation are different from the unrestricted coefficients, and the goodness-of-fit test against the unrestricted model is highly significant.

The point estimates imply that mechanics' employment responds negatively to increases in their alternative wage rate, as measured by the manufacturing wage rate, and positively to increases in their contractual wage rate. The latter effect, which may be taken as weak evidence for a positive correlation between contract wages and employment, is consistent with either the expected utility preference specification (2) and some degree of risk aversion, or the Cobb-Douglas specification (3) with $\theta < 1$. From equation (13), the elasticity of employment with respect to the contract wage is $b_2 = \alpha\eta$ and the elasticity of contract employment with respect to the alternative wage is $b_3 = (1-\alpha)\eta$, where $\eta = -\gamma_2/(\gamma_1+\gamma_2)$ is the constant-output employment elasticity associated with the Cobb-Douglas maintenance technology, and α is the relative weight of contract wages in the expression for the shadow value of labor. The point estimates in column 5 of Table 8 imply $\alpha = -.59$ and $\eta = .-64$, although the estimates are extremely imprecise and insignificantly different from zero. For the expected utility specification of workers' preferences, the coefficient $-\alpha$ represents the product of the relative risk-aversion coefficient (δ) and the average markup of contract wages over alternative wages (π). If the latter is about .25, the implied estimate of the relative risk-aversion coefficient is about 2.4. For the Cobb-Douglas preference specification, the coefficient α is an estimate of $(1-(1/\theta))$. The implied value of θ is .63. These estimates are not unreasonable, although their imprecision is disturbing, as is the failure of the reduced form of the model to reproduce the unrestricted reduced form of the data.

---

[34] When prices were excluded from the forecasting equation for contract wages, the sum-of-square function contained a very flat ridge in the $b_2 - b_3$ plane. The fit of the model was essentially unchanged with b_2 large and negative and b_3 large and positive, so long as their ratio was approximately $-.60$.

Overall, none of the models considered in this paper gives a particularly good fit to the data, particularly with respect to the coefficients of contract or manufacturing wages. In addition, the fact that real contract wage rates and real manufacturing wage rates are both heavily influenced by lagged consumer prices is not easily reconciled with the absence of price effects on employment.[35] A more flexible model of the interactions between contract and alternative wage rates, on one hand, and employment, on the other, is apparently needed to describe the data.

IV. Conclusions

This paper presents an analysis of firm-specific employment and wage outcomes for airline mechanics at seven firms during the period 1969–76. The data possess many of the familiar properties of aggregate wage and employment data, including second-order serial correlation in employment and first-order serial correlation in real wages. Airline mechanics' employment levels are found to be correlated with both their own wage rates, and with average wage rates outside the airline industry.

A theoretical model is presented that describes the evolution of wages and employment in several alternative settings, including the traditional labor demand setting, where firms take contract wages as exogenous, and an efficient contract setting, where wages and employment are jointly determined to minimize employer costs, subject to a utility requirement for workers. The model incorporates costly adjustment of employment over time and emphasizes that workers' alternative wages can have two effects on employment outcomes: a direct effect on the shadow value of workers' time, and an indirect effect on forecasts of future wage outcomes.

The model gives a straightforward and relatively successful interpretation of the empirical link between airline departures and mechanics' employment. None of the alternative versions of the model, however, successfully captures the links between wages and employment. Both the labor demand model and the simplest efficient contracting model, which equates the marginal product of workers to their alternative wage rate, are rejected in favor of a more general model that includes contract and alternative wages in the employment equation. The parameter estimates for this model, however, are extremely imprecise, and the implied reduced-form employment equation fits poorly relative to an unrestricted autoregression. In spite of the promise that simple contracting models might provide a credible interpretation of observed movements in employment and wages, the covariation of employment and wages in these data remains largely unexplained.

APPENDIX

I. *Derivation of the Reduced-Form Employment Equation*

Let $y_t' = (\log F_t, \log F_{t-1}, \log w_t, \log w_{t-1}, \log a_t, \log a_{t-1}, \log p_t, \log p_{t-1})$ represent the vector of current and once-lagged values of departures, contract wages, manufacturing wages, and prices. Equation (8) can be written as $y_t = Ay_{t-1} + u_t$, where $u_t = (u_{1t}, 0, u_{2t}, 0, u_{3t}, 0, u_{4t}, 0)$, and A is a suitably defined matrix of coefficients. Let $e_1' = (1,0,\dots,0)$. Then

$$(1-\lambda\beta)\sum_{j=0}^{\infty}(\lambda\beta)^j E_t \log F_{t+j}$$

$$= e_1'(1-\lambda\beta)\sum_{j=0}^{\infty}(\lambda\beta A)^j y_t,$$

with similar expressions for the forward-moving averages of contract wages and manufacturing wages. Provided that the characteristic roots of A are smaller than $(\lambda\beta)^{-1}$ in modulus, the infinite sum $\Sigma_{j=0}^{\infty}(\lambda\beta A)^j$ converges. Assuming this to be true, let

---

[35] Brown and Ashenfelter report similar findings in their investigation of wage and employment outcomes for typographers.

$$(1-\lambda\beta)\sum_{j=0}^{\infty}(\lambda\beta A)^j = A^*.$$

According to equation (16),

$$\text{(A1)}\quad \log N_t = \lambda \log N_{t-1} + (1-\lambda) \times [b_1 e_1' A^* + b_2 e_3' A^* + b_3 e_5' A^*] y_t + \varepsilon_t^*,$$

where ε_t^* is first-order autoregressive with autoregressive parameter ρ. Let

$$a^* = b_1 e_1' A^* + b_2 e_3' A^* + b_3 e_5' A^*.$$

Performing a transformation of (A1) to eliminate the serial correlation in ε_t^* yields

$$\text{(A2)}\quad \log N_t = (\lambda+\rho)\log N_{t-1} - \lambda\rho \log N_{t-2} + (1-\lambda)a^* y_t - \rho(1-\lambda)a^* y_{t-1} + \xi_t,$$

where ξ_t is serially uncorrelated. Finally, substituting for y_t into (A2) yields

$$\text{(A3)}\quad \log N_t = (\lambda+\rho)\log N_{t-1} - \lambda\rho \log N_{t-2} + [(1-\lambda)a^* A - \rho(1-\lambda)a^*] y_{t-1} + \xi_t^*,$$

where $\xi_t^* = \xi_t + (1-\lambda)a^* u_t$ is serially uncorrelated. In this paper I compute the matrix A^* numerically and use the resulting estimates to compute the coefficients in (A3) for each value of $\lambda, \rho, b_1, b_2, b_3$, and the forecasting coefficients in A.

II. *Data Definitions and Sources*

1. Wages

a) Wage rate information for airline mechanics was assembled from three sources: contract listings in the Bureau of Labor Statistics' *Current Wage Developments* (*CWD*), contract listings and reports in the Bureau of National Affairs' *Daily Labor Report* (*DLR*), and copies of the collective bargaining agreements. Typically, contract listings in *CWD* and *DLR* include information on percentage and/or absolute wage increases for certified mechanics in each year of the contract. *CWD* also lists cost-of-living adjustments as they accrue over the life of the contract. Information on wage changes was combined with sporadic information on the level of wage rates from both sources to produce a time-series of wage rates by quarter for each airline. For mechanics at Western Airlines, I constructed a time-series of wage rates directly from a complete contract chronology generously provided by the International Brotherhood of Teamsters Airline Division. Wage rates were calculated up to 1984, and compared with the rates listed in recent labor contracts for each airline on file at the National Mediation Board in Washington, D.C. In cases where a wage rate was increased retroactively, I use the wage rate that actually prevailed, rather than the retroactive wage rate. In cases where a wage rate increased within a quarter (say in February), I use the unweighted average of the beginning and ending rates as an average quarterly rate.

(*i*) American wages are benchmarked to a September 1982 rate of 13.53 reported in *DLR* (March 7, 1983). New contracts were negotiated March 1969 (*CWD*, April 1969), May 1971 (*CWD*, July 1971), May 1974 (*CWD*, May 1974), and June 1975 (*CWD*, August 1975).

(*ii*) Braniff wages are benchmarked to a rate of 4.14 (excluding COLA) in January 1969 reported in *DLR* (May 6, 1969). New contracts were negotiated May 1969 (*DLR*, May 6, 1969), September 1972 (*CWD*, November 1972), November 1973 (*CWD*, January 1974) and March 1976 (*CWD*, May 1976).

(*iii*) Continental wages are benchmarked to a rate of 4.14 (excluding COLA) in January 1969. New contracts were negotiated September 1969 (*CWD*, October 1969), August 1972, August 1974 (*CWD*, September 1974) and January 1976 (*CWD*, March 1976).

(*iv*) Eastern wage rates are benchmarked to a rate of 7.69 (excluding COLA) in November 1975 reported in *DLR* (July 12, 1974). New contracts were negotiated May 1969 (*DLR*, May 2, 1969). October 1972 (*CWD*, January 1973, and later Payboard rulings in *CWD*, May 1973), July 1974 (*DLR*, July 12, 1974), and January 1976 (*CWD*, March 1976).

(*v*) Trans World wage rates are benchmarked to a rate of 7.67 (excluding COLA) in May 1975 reported in *DLR* (March 13, 1975). New contracts were negotiated January 1970 (*CWD*, February 1970), December 1972 (*CWD*, February 1973), February 1975 (*DLR*, March 13, 1975), and September 1976 (*CWD*, November 1976).

(*vi*) United wage rates are benchmarked to a rate of 7.50 (excluding COLA) in January 1975, reported in

DLR (May 3, 1974). New contracts were negotiated October 1969 (*CWD*, November 1969), August 1972 (*CWD*, October 1972), May 1974 (*CWD*, June 1974) and December 1975 (*CWD*, February 1976).

(*vii*) Western wage rates are taken directly from collective bargaining agreements dated November 1968 (negotiated August 1969); dated November 1971 (negotiated July 1972); and dated November 1973 (negotiated May 1974). Information on COLA payments under the Western agreements was taken from *CWD* (various issues).

b) Wage rate information for maintenance mechanics in manufacturing industries is taken from the Bureau of Labor Statistics' *Area Wage Survey*. Results of the annual survey, which is administered in different months in different cities, are published in the following BLS *Bulletins*: 1968–No. 1575; 1969–No. 1625; 1970–No. 1660; 1971–No. 1685; 1972–No. 1725; 1973–No. 1775; 1974–No. 1850; 1975–No. 1900; 1976–No. 1950 (Washington: USGPO). I used both the average wage rate for all cities, and wage rates for the following city-airline pairs: Dallas/Fort Worth–Braniff (weighted average of Dallas (.67) and Fort Worth (.33) from 1968 to 1973); Miami–Eastern; Oklahoma City (in lieu of Tulsa)–American; St. Louis–Trans World; Los Angeles–Continental and Western; San Francisco–United. The annual wage series were interpolated linearly using the 1968 values for first quarter 1969 for cities where the survey is administered in late fall (Dallas/Fort Worth, Miami, Oklahoma City, San Francisco) and using the 1969 values for first quarter 1969 for St. Louis and Los Angeles, where the surveys are administered earlier in the year. The average wage rate for all cities, which the BLS reports as an average for February, was interpolated using the 1969 wage for first quarter 1969.

c) Average straight-time hourly earnings of production workers in manufacturing represent quarterly averages of monthly earnings as reported in U.S. Bureau of Commerce, *Business Statistics* (1979 ed.) (Washington: USGPO, 1980, p. 228).

d) Wage rates of licensed mechanics at Boeing are taken from the BLS Wage Chronology for Boeing (Washington State) and the International Association of Machinists, published in BLS *Bulletin* No. 1895 (Washington: USGPO, 1976). Rates for mechanics are listed as basic hourly rates for labor grade 10 in Table 2c, p. 19.

2. Flight Information

Flight information is taken from monthly information in Civil Aeronautics Board, *Air Carrier Traffic Statistics* (Washington: CAB).

a) Departures: Aircraft revenue departures performed in each quarter in scheduled domestic operations.

b) Available Seat Miles: Total available seat miles in each quarter in scheduled domestic operations.

c) Revenue Passenger Miles: Total revenue passenger miles in each quarter in scheduled domestic operations.

d) Domestic Flight Hours: Total aircraft revenue hours in scheduled domestic operations and nonscheduled domestic operations.

e) Total Flight Hours: Total aircraft revenue hours in all services (domestic and international; scheduled and nonscheduled).

Precise definitions of these variables are contained in each issue of *Air Carrier Traffic Statistics*.

3. Employment Information

Employment and payroll information is taken from the "Form-41" reports filed monthly by each airline. Prior to 1977, the March, June, September, and December Form 41 reports contained a quarterly P–10 report on employment and payroll by category of worker. The number of mechanics is taken from employment totals for maintenance labor (P–10 Account Number 25). In Figure 1, "pilots" include pilots and copilots (P–10 Account Number 23) and flight engineers (P–10 Account Number 5124); "flight attendants" represent passenger service flying personnel (P–10 Account Number 5524). Form 41 reports are on file at the Department of Transportation in Washington D.C. (Formerly, these reports were housed at the Civil Aeronautics Board.) The 1969 information was taken from microfilm archives of the Form 41 reports.

4. Miscellaneous Information

a) Real GNP (seasonally adjusted) is taken from U.S. Bureau of Commerce *Business Statistics* (1979 ed.) (Washington: USGPO, 1980, p. 251).

b) The Consumer Price Index (All Items) is taken from OECD, *Historical Statistics 1960–79* (Paris: OECD, 1980, p. 73).

c) A price index for domestic passenger fares was constructed from information contained in *Annual Reports* of various airline firms, and from the Civil Aeronautics Board *Annual Report* (Washington: CAB), various issues. Between January 1969 and December 1976 the major changes in fares were as follows:

February 20, 1969: 3.8 percent general increase
October 1, 1969: 6.35 percent general increase
October 15, 1970: 3.0 percent general increase
May 1, 1971: 6.0 percent general increase
September 5, 1972: 2.7 percent general increase
December 1, 1973: 5.0 percent general increase
November 1, 1974: 4.0 percent general increase
November 15, 1975: 3.0 percent general increase.

The Civil Aeronautics Board also granted an 8.0 percent general fare increase in a series of three increments during 1976. I assume that these increases took the form of a 2.0 percent increase on April 1, 1976, a 3.0 percent general increase on July 1, 1976, and a 2.0 percent general increase on October 1, 1976.

d) An index of jet fuel prices was constructed from annual average jet fuel prices per gallon reported in Civil Aeronautics Board, "Long Term Fuel Expense, Consumption, Unit Prices and Total Expense, System Trunks and Locals," CAB Financial and Cost Analysis Division (Manuscript dated March 21, 1983), index-linked within each year to quarterly averages of monthly producer prices for refined petroleum products, as reported in OECD, *Historical Statistics 1960–79* (Paris: OECD, 1980, p. 73). The CAB reports average annual fuel prices (in cents per gallon) as 12.23 in 1972, 12.85 in 1973, 21.95 in 1974, 27.37 in 1975, 30.42 in 1976, and

35.24 in 1977. I used the producer price index for refined petroleum, indexed-linked to the 1972 average price reported by the CAB, to estimate quarterly prices for jet fuel in 1969, 1970, and 1971. Fuel prices reported by the various individual airlines in their *Annual Reports* are very similar to the CAB average prices. For example, in 1977, Western reported an average price of .36, United reported an average price of .35, Trans World reported an average price of .375 and Continental reported an average price of .359 per gallon.

e) An index of aircraft parts prices was constructed from the monthly industry wholesale price index for "General Purpose Machinery and Parts" (Commodity Group 114) from Bureau of Labor Statistics *Wholesale Prices and Price Indexes* (Washington: USGPO), various issues.

f) Fleet composition information by airline was interpolated from annual fleet composition information contained in the Federal Aviation Authority *Statistical Handbook of Aviation* (Washington: FAA), various issues.

g) Airline-specific unemployment rates are based on quarterly averages of monthly insured unemployment rates for the following city-airline pairs: Los Angeles–Continental and Western; San Francisco–United; Miami–Eastern; St. Louis–Trans World; Dallas/Fort Worth–Braniff. I used the insured unemployment rate for Oklahoma in lieu of city-specific information for Tulsa–American. Insured unemployment rates by city and state were taken from BLS *Employment and Earnings* (Washington: USGPO) various issues.

A complete listing of the data is available from the author on request. Copies of the computer programs used to generate the estimates in Tables 7 and 8 are also available on request.

REFERENCES

Ashenfelter, Orley and Card, David, "Time Series Representations of Economic Variables and Alternative Models of the Labor Market," *Review of Economic Studies*, Special Issue 1982, *49*, 761–82.

Azariadis, Costas, "Implicit Contracts and Underemployment Equilibria," *Journal of Political Economy*, December 1975, *83*, 1183–202.

Baily, Martin N., "Wages and Employment Under Uncertain Demand," *Review of Economic Studies*, January 1974, *41*, 37–50.

Barro, Robert, "Long Term Contracting, Sticky Prices, and Monetary Policy," *Journal of Monetary Economics*, July 1977, *3*, 306–16.

Brown, James N. and Ashenfelter, Orley, "Testing the Efficiency of Employment Contracts," *Journal of Political Economy*, June 1986, *94*, S40–S87.

Carruth, Alan A. and Oswald, Andrew, "Miners Wages in Post-War Britain: An Application of a Model of Trade Union Behavior," *Economic Journal*, December 1985, *95*, 1003–20.

Gallant, A. Ronald and Jorgenson, D. W., "Statistical Inference for a System of Simultaneous, Non-linear, Implicit Equations in the Context of Instrumental Variables Estimation," *Journal of Econometrics*, October/December 1979, *11*, 275–302.

Gordon, Donald F., "A Neo-Classical Theory of Keynesian Unemployment," *Economic Inquiry*, December 1974, *12*, 431–59.

Hall, Robert E. and Lilien, David, "Efficient Wage Bargains Under Uncertain Supply and Demand," *American Economic Review*, December 1979, *69*, 868–79.

Hart, Oliver, "Optimal Labor Contracts Under Asymmetric Information: An Introduction," *Review of Economic Studies*, January 1983, *50*, 3–35.

Kennan, John, "The Estimation of Partial Adjustment Models with Rational Expectations," *Econometrica*, November 1979, *47*, 1441–56.

Leontief, Wassily, "The Pure Theory of the Guaranteed Annual Wage Contract," *Journal of Policy Economy*, February 1946, *54*, 76–79.

McDonald, Ian M. and Solow, Robert, "Wage Bargaining and Employment," *American Economic Review*, December 1981, *71*, 896–908.

MaCurdy, Thomas and Pencavel, John, "Testing Between Competing Models of Wage and Employment Determination in Unionized Labor Markets," *Journal of Political Economy*, June 1986, *94*, S3–S39.

Martinello, F., "Wage and Employment Determination in a Unionized Industry: The IWA in the B. C. Wood Products Industry," Economics Paper 84–90, Carleton University, March 1984.

Pencavel, John H., "The Trade-off Between Wages and Employment in Trade Union Objectives," *Quarterly Journal of Econom-*

ics, May 1984, *99* 215–32.

Rosen, Sherwin, "Implicit Contracts—A Survey," *Journal of Economic Literature*, September 1985, *23*, 1144–75.

Sargent, Thomas J., "Estimation of Dynamic Labor Demand Schedules Under Rational Expectations," *Journal of Political Economy*, December 1978, *86*, 1009–44.

Svejnar, Jan, "Bargaining Power, Fear of Disagreement and Wage Settlements: Theory and Evidence from U.S. Industry," *Econometrica*, forthcoming 1986.

Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show

By WILLIAM J. BAUMOL*

Maddison's 1870–1979 data are analyzed, showing the historically unprecedented growth in productivity, gross domestic product per capita and exports and the remarkable convergence of productivities of industrialized market economies, with convergence apparently shared by planned economies but not less developed countries. Productivity lag's relation to "deindustrialization," unemployment, and balance of payments is examined. The data are shown to suggest a tempered view of the slowdown in U.S. productivity growth and its lag behind other countries.

> No matter how refined and how elaborate the analysis, if it rests solely on the short view it will still be...a structure built on shifting sands.
>
> Jacob Viner [1958, pp. 112–13]

Recent years have witnessed a reemergence of interest on the part of economists and the general public in issues relating to long-run economic growth. There has been a recurrence of doubts and fears for the future—aroused in this case by the protracted slowdown in productivity growth since the late 1960's, the seeming erosion of the competitiveness of U.S. industries in world markets, and the spectre of "deindustrialization" and massive structural unemployment. These anxieties have succeeded in redirecting attention to long-run supply-side phenomena that formerly were a central preoccupation of economists in the industrializing West, before being pushed aside in the crisis of the Great Depression and the ensuing triumph of Keynesian ideas.

Anxiety may compel attention, but it is not necessarily an aid to clear thinking. For all the interest now expressed in the subject of long-run economic growth and policies ostensibly directed to its stimulation, it does not seem to be widely recognized that adequate economic analysis of such issues calls for the careful study of economic history—if only because it is there that the pertinent evidence is to be found. Economic historians have provided the necessary materials, in the form of brilliant insights, powerful analysis, as well as a surprising profusion of long-period data. Yet none of these has received the full measure of attention they deserve from members of the economics profession at large.

To dramatize the sort of reorientation long-term information can suggest, imagine a convincing prediction that over the next century, U.S. productivity growth will permit a trebling of per capita GNP while cutting nearly by half the number of hours in the average work year, and that this will be accompanied by a sevenfold increase in exports. One might well consider this a very rosy forecast. But none of these figures is fictitious. For these developments in fact lay before the United Kingdom in 1870, just as its economic leadership began to erode.

---

*Princeton University, Princeton, NJ 08544, and New York University. I express my deep gratitude to the Division of Information Science and Technology of the National Science Foundation, the Exxon Education Foundation, the Joint Council on Economic Education, the Fishman–Davidson Center for the Study of the Service Sector, and the C. V. Starr Center for Applied Economics for their generous support of the research underlying this paper. I am also heavily indebted to Edward Wolff, Wayne Farel, Robert Dorfman, Sidney Ratner, and Mariza Stipec for help in various parts of the work. Valuable suggestions from several anonymous reviewers helped to limit the errors of this amateur economic historian. I also learned a great deal from the excellent Abramovitz paper (1985) on the subject. Above all, I owe an enormous debt to Paul David for his encouragement, and for perceptive and helpful comments which guided the final revision of this paper. Entire paragraphs come from his pen.

This paper outlines some implications of the available long-period data on productivity and related variables—some tentative, some previously noted by economic historians, and some throwing a somewhat surprising light on developments among industrialized nations since World War II. Among the main observations that will emerge here is the remarkable convergence of output per labor hour among industrialized nations. Almost all of the leading free enterprise economies have moved closer to the leader, and there is a strong inverse correlation between a country's productivity standing in 1870 and its average rate of productivity growth since then. Postwar data suggest that the convergence phenomenon also extends to both "intermediate" and centrally planned economies. Only the poorer less developed countries show no such trend.

It will also emerge that over the century, the U.S. productivity growth rate has been surprisingly steady, and despite frequently expressed fears, there is no sign recently of any *long-term* slowdown in growth of either total factor productivity or labor productivity in the United States. And while, except in wartime, *for the better part of a century*, U.S. productivity growth rates have been low relative to those of Germany, Japan, and a number of other countries, this may be no more than a manifestation of the convergence phenomenon, which requires countries that were previously behind to grow more rapidly. Thus, the paper will seek to dispel these and a number of other misapprehensions apparently widespread among those who have not studied economic history.

Nonspecialists may well be surprised at the remarkably long periods spanned in time-series contributed by Beveridge, Deane, Kuznets, Gallman, Kendrick, Abramovitz, David, and others. The Phelps Brown-Hopkins indices of prices and real wages extend over seven centuries. Maddison, Feinstein (and his colleagues), and Kendrick cover productivity, investment, and a number of other crucial variables for more than 100 years. Obviously, the magnitudes of the earlier figures are more than a little questionable, as their compilers never cease to warn us. Yet the general qualitative character of the time paths are persuasive, given the broad consistency of the statistics, their apparent internal logic and the care exercised in collecting them. In this paper, the period used will vary with topic and data availability. In most cases, something near a century will be examined, using primarily data provided by Angus Maddison (1982) and R. C. O. Matthews, C. H. Feinstein, and J. C. Odling-Smee (1982—henceforth, M-F-O).[1]

I. The Magnitude of the Accomplishment

The magnitude of the productivity achievement of the past 150 years resists intuitive grasp, and contrasts sharply with the preceding centuries. As the *Communist Manifesto* put the matter in 1848, with remarkable foresight, "The bourgeoisie, during its rule of scarce one hundred years, has created more massive and more colossal productive forces than have all preceding generations together." There obviously are no reliable measures of productivity in antiquity, but available descriptions of living standards in Ancient Rome suggest that they were in many respects higher than in eighteenth-century England (see Colin Clark, 1957, p. 677). This is probably true even for the lower classes—certainly for the free urban proletariat, and perhaps even with the inclusion of slaves. An upper-class household was served by sophisticated devices for heating and bathing not found in eighteenth-century homes of the rich. A wealthy Roman magically transported into an eighteenth-century English home would probably have been puzzled by the technology of only a few products—clocks, window panes, printed books and newspapers, and the musket over the fireplace.

[1] The Maddison absolute productivity figures will be used in preference to the M-F-O data, since the former include more years and more countries. However, the M-F-O series has one advantage. They report productivity statistics only for years which can be considered to contain peaks of business cycles, so that the calculations are not distorted by the well-known effects of the business cycle on labor productivity. Yet over the long period in question here, no remarkable differences in patterns seem to emerge from the two sets of figures.

It is true that even during the Middle Ages (see, for example, Carlo Cipolla, 1976), there was substantial technological change in the workplace and elsewhere. Ship design improved greatly. Lenses and, with them, the telescope and microscope appeared in the sixteenth century, and the eighteenth century brought the ship's chronometer which revolutionalized water transport by permitting calculation of longitude. Yet, none of this led to rates of productivity growth anywhere near those of the nineteenth and twentieth centuries.

Nonhistorians do not usually recognize that initially the Industrial Revolution was a fairly minor affair for the economy as a whole. At first, much of the new equipment was confined to textile production (though some progress in fields such as iron making had also occurred). And, as David Landes (1969) indicates,[2] an entrepreneur could undertake the new types of textile operations with little capital, perhaps only a few hundred pounds, which (using the Phelps Brown-Hopkins data) translates into some 100,000 1980 dollars. Jeffrey Williamson (1983) tells us that in England during the first half-century of the Industrial Revolution, real per capita income grew only about 0.3 percent per annum,[3] in contrast with the nearly 3 percent achieved in the Third World in the 1970's (despite the decade's economic crises).

TABLE 1—TOTAL GROWTH FROM 1870 TO 1979[a]
PRODUCTIVITY, *GDP* PER CAPITA, AND EXPORTS
SIXTEEN INDUSTRIALIZED COUNTRIES[b]

	Real *GDP* per Work-Hour	Real *GDP* per Capita	Volume of Exports
Australia	398	221	–
United Kingdom	585	310	930
Switzerland	830	471	4,400
Belgium	887	439	6,250
Netherlands	910	429	8,040
Canada	1,050	766	9,860
United States	1,080	693	9,240
Denmark	1,098	684	6,750
Italy	1,225	503	6,210
Austria	1,270	643	4,740
Germany	1,510	824	3,730
Norway	1,560	873	7,740
France	1,590	694	4,140
Finland	1,710	1,016	6,240
Sweden	2,060	1,083	5,070
Japan	2,480	1,661	293,060

Source: Angus Maddison (1982, pp. 8, 212, 248–53).
[a] In 1970 U.S. dollars.
[b] Shown in percent.

Table 1 shows the remarkable contrast of developments since 1870 for Maddison's 16 countries. We see (col. 1) that growth in output per work-hour ranged for the next 110 years from approximately 400 percent for Australia all the way to 2500 percent (in the case of Japan). The 1100 percent increase of labor productivity in the United States placed it somewhat below the middle of the group, and even the United Kingdom managed a 600 percent rise. Thus, after not manifesting any substantial long-period increase for at least 15 centuries, in the course of 11 decades the median increase in productivity among the 16 industrialized leaders in Maddison's sample was about 1150 percent. The rise in productivity was sufficient to permit output per capita (col. 2) to increase more than 300 percent in the United Kingdom, 800 percent in West Germany, 1700 percent in Japan, and nearly 700 percent in France and the United States. Using Robert Summers and Alan Heston's sophisticated international comparison data (1984), this implies that in 1870, U.S. output per capita was comparable to 1980 output per capita in Honduras and the Philippines, and slightly below that of China, Bolivia, and Egypt!

---

[2] "The early machines, complicated though they were to contemporaries, were nevertheless modest, rudimentary, wooden contrivances which could be built for surprisingly small sums. A forty-spindle jenny cost perhaps £6 in 1772; scrubbing and carding machines cost £1 for each inch of roller width; a clubbing billy with thirty spindles cost £10.10s" (Landes, pp. 64–65). This suggests at least the possibility (pointed out by Landes) that part of the reason investment was low is that not very much capital may have been required.

[3] This observation does not quite seem to square with Charles Feinstein's estimates (1972, pp. 82–94) which indicate that while output per worker in the United Kingdom increased 0.2 percent per year between 1761 and 1800, between 1801 and 1830 the growth rate leaped up to 1.4 percent per annum. He estimates that total factor productivity behaved similarly. However, between 1801 and 1810, total annual investment fell to 10 percent of gross domestic product, in comparison with its 14 percent rate in the immediately preceding and succeeding periods.

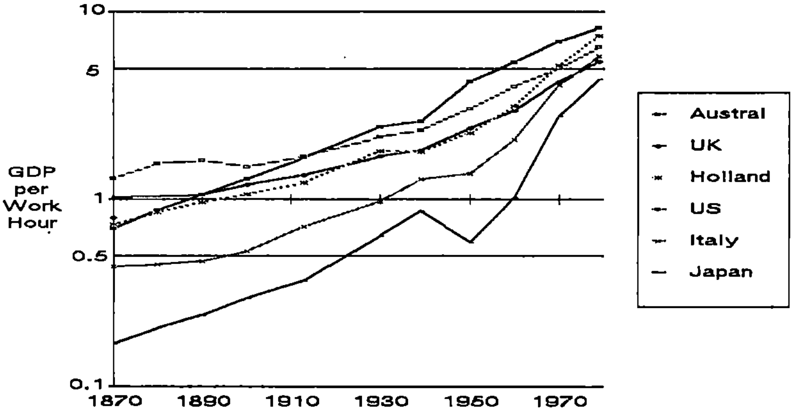


FIGURE 1. GROSS DOMESTIC PRODUCT PER WORK-HOUR, 1870–1979

Source: Maddison (p. 212)

The growth rates of other pertinent variables were also remarkable. One more example will suffice to show this. Table 1, which also shows the rise in volume of exports from 1870 to 1979 (col. 3) indicates that the median increase was over 6000 percent.

II. The Convergence of National Productivity Levels

There is a long and reasonably illustrious tradition among economic historians centered on the phenomenon of convergence. While the literature devoted to the subject is complex and multifaceted, as revealed by the recent reconsideration of these ideas by Moses Abramovitz (1985), one central theme is that forces accelerating the growth of nations who were latecomers to industrialization and economic development give rise to a long-run tendency towards convergence of levels of per capita product or, alternatively, of per worker product. Such ideas found expression in the works of Alexander Gerschenkron (see, for example, 1952), who saw his own views on the advantages of "relative backwardness" as having been anticipated in important respects by Thorstein Veblen's writings on the penalties of being the industrial leader (1915). Although such propositions also have been challenged and qualified (for example, Edward Ames and Nathan Rosenberg, 1963), it is difficult to dismiss the idea of convergence on the basis of the historical experience of the industrialized world. (For more recent discussions, see also the paper by Robin Marris, with comments by Feinstein and Matthews in Matthews, 1982, pp. 12–13, 128–147; as well as Dennis Mueller, 1983.)

Using 1870–1973 data on gross domestic product (*GDP*) per work-year for 7 industrialized countries, M-F-O have shown graphically that those nations' productivity levels have tended to approach ever closer to one another. The same phenomenon for 6 countries is illustrated in Figure 1 in a semilog representation based on Maddison's data for 1870–1979, which provide estimates of output per work-hour for 16 countries.[4]

The convergence toward the vanguard (led in the first decades by Australia—see Richard Caves and Laurence Krause, 1984 —and the United Kingdom and, approximately since World War I, by the United

---

[4] Space prevents extensive consideration of Paul Romer's (1985) objection to the evidence offered for the convergence hypothesis provided here and elsewhere, i.e., that the sample of countries studied is an *ex post* selection of successful economies. Successes, by definition, are those which have done best relative to the leader. However, the Summers-Heston 1950–80 data for 72 countries represented in Figure 3 do permit an *ex ante* selection. Tests ranking countries both by 1950 and by 1960 *GDP* levels confirm that even an *ex ante* sample of the wealthiest countries yields a pattern of convergence which, while less pronounced than that calculated from an *ex post* group, is still unambiguous.

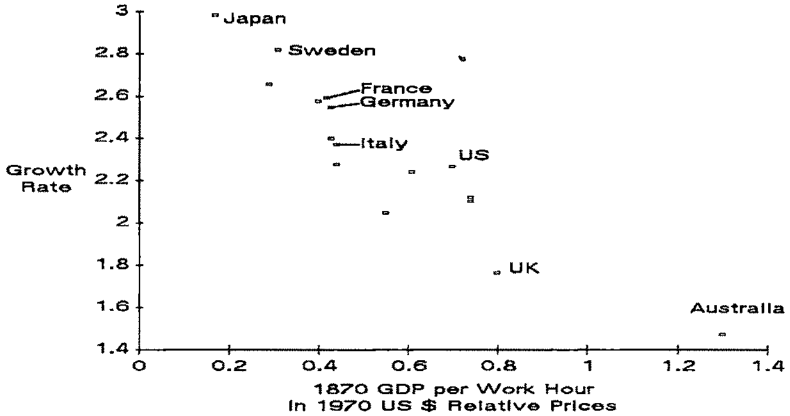


FIGURE 2. PRODUCTIVITY GROWTH RATE, 1870–1979 VS. 1870 LEVEL

Source: Maddison (p. 212)

States) is sharper than it may appear to the naked eye. In 1870, the ratio of output per work-hour in Australia, then the leader in Maddison's sample, was about eight times as great as Japan's (the laggard). By 1979, that ratio for the leader (the United States) to the laggard (still Japan) had fallen to about 2. The ratio of the standard deviation from the mean of *GDP* per work-hour for the 16 countries has also fallen quite steadily, except for a brief but sharp rise during World War II.

The convergence phenomenon and its pervasiveness is confirmed by Figure 2, on which my discussion will focus. The horizontal axis indicates each Maddison country's absolute level of *GDP* per work-hour in 1870. The vertical axis represents the growth rate of *GDP* per work-hour in the 110 years since 1870. The high inverse correlation between the two is evident. Indeed, we obtain an equation (subject to all sorts of statistical reservations)[5]

$$\textit{Growth Rate}\ (1870\text{–}1979)$$
$$= 5.25 - 0.75\ln\,(\textit{GDP}\ \text{per}\ \textit{WorkHr},\ 1870),$$
$$R^2 = 0.88.$$

That is, with a very high correlation coefficient, the higher a country's productivity level in 1870 the more slowly that level grew in the following century.

---

[5] The high correlation should not be taken too seriously. Aside from the reasons why its explanation may be misunderstood that are presently discussed in the text, the tight fit of the data points is undoubtedly ascribable in good part to several biassing features of the underlying calculation. First, the 1870 figures were calculated by Maddison using backward extrapolation of growth rates, and hence their correlation is hardly surprising. Second, since growth rate, r, is calculated by solving $y_t = e^{rt}y_0$ for r, to obtain $r = (\ln y_t - \ln y_0)/t$, where $y_t = GDP$ per capita in period t, a regression equation $r = f(y_0)$ contains the same variable, y_0 on both sides of the equation, thus tending to produce a spurious appearance of close relationship. Indeed, if the convergence process were perfect, so that we would have $y_t = k$ with k the same for every country in the sample, every dot in the diagram would necessarily perfectly fit the curve $r = lnk/t - lny_0/t$, and the r^2 would be unity, identically. As we will see, however, the 72-country data depicted in Figure 3 hardly constitute a close fit (the R^2 is virtually zero), and do not even yield a negatively sloping regression line. Thus, a relationship such as that in Figure 2 is no tautology, nor even a foregone conclusion.

In addition, if the 1870 productivity levels are measured with considerable error, this must result in some significant downward bias in the regression coefficient on $ln(GDP$ per $WorkHr$, 1870$)$. This is a point distinct from the one concerning the size of the correlation coefficient, although the latter is affected by the fact that relatively large measurement errors in the 1870 productivity levels enter as inversely correlated measurement errors in the 1870–1979 growth rate. The argument that this bias is not sufficient to induce a negative correlation in the 72-country sample may not be wholly germane, as the relative seriousness of the measurement errors in the initial and terminal observations may be much the same for observations confined to the period 1950–80.

TABLE 2—RELATIVE GROWTH IN REAL WAGES, *GDP* PER WORK-HOUR AND CAPITAL STOCK, UNITED KINGDOM AND GERMANY, 1860–1980

	Period	Ratio: German Increase to U.K. Increase[b]
Real Wages	1860–1980	4.25
GDP per Labor Hour	1870–1979	2.35
Capital Stock[a]	1870–1979	6.26
Capital Stock per Worker	1870–1979	3.8
Capital Stock per Capita	1870–1979	5.4

Sources: Real wages, same as in fn. 6; all other data from Maddison.

[a] Net nonresidential fixed tangible capital stock.

[b] (German 1979 figure/German 1870 figure)/(U.K. 1979 figure/U.K. 1870 figure) with appropriate modification of the dates for the wage figures.

III. Implications of the Inverse Correlation: Public Goods Property of Productivity Policy

The strong inverse correlation between the 1870 productivity levels of the 16 nations and their subsequent productivity growth record seems to have a startling implication. Of course, hindsight always permits "forecasts" of great accuracy—that itself is not surprising. Rather, what is striking is the apparent implication that *only one variable*, a country's 1870 *GDP* per work-hour, or its relation to that of the productivity leader, matters to any substantial degree, and that other variables have only a peripheral influence. It seems not to have mattered much whether or not a particular country had free markets, a high propensity to invest, or used policy to stimulate growth. Whatever its behavior, that nation was apparently fated to land close to its predestined position in Figure 2.

However, a plausible alternative interpretation is that while national policies and behavior patterns do substantially affect productivity growth, the spillovers from leader economies to followers are large—at least among the group of industrial nations. If country *A*'s extraordinary investment level and superior record of innovation enhances its own productivity, it will almost automatically do the same in the long run for industrialized country *B*, though perhaps to a somewhat more limited extent. In other words, for such nations a successful productivity-enhancing measure has the nature of a public good. And because the fruits of each industrialized country's productivity-enhancement efforts are ultimately shared by others, each country remains in what appears to be its predestined *relative* place along the growth curve of Figure 2. I will note later some considerations which might lead one to doubt that the less developed countries will benefit comparably from this sharing process.

This sharing of productivity growth benefits by industrialized countries involves both innovation and investment. The innovation-sharing process is straightforward. If industry in country *A* benefits from a significant innovation, those industries in other countries which produce competing products will find themselves under pressure to obtain access to the innovation, or to an imitation or to some other substitute. Industrialized countries, whose product lines overlap substantially and which sell a good deal in markets where foreign producers of similar items are also present, will find themselves constantly running in this Schumpeterian race, while those less developed countries which supply few products competing with those of the industrialized economies will not participate to the same degree.

There is reason to suspect that the pressures for rapidity in imitation of innovation in industrial countries have been growing. The explosion in exports reported in Table 1 has given them a considerably larger share of gross national product than they had in 1870. This suggests that more of each nation's output faces the direct competition of foreign rivals. Thus, the penalties for failure to keep abreast of innovations *in other countries* and to imitate them where appropriate have grown.

Second, the means required for successful imitations have improved and expanded enormously. World communications are now practically instantaneous, but required weeks and even months at the birth of the Industrial Revolution. While today meetings of scientists and technicians are widely encouraged, earlier mercantilistic practices entailed measures by each country to prevent other nations from learning its industrial techniques, and the emigration of specialized workers was often forbidden. Though figures in this arena are difficult to interpret, much less substantiate, one estimate claims that employment in "information activities" in the United States has grown from less than 1 percent of the labor force in 1830 to some 45 percent today (James Beniger, forthcoming, p. 364, leaning heavily on Marc Porat, 1977). Presumably, growth of the information sector in other industrialized nations has been similar. This must surely facilitate and speed the innovative, counterinnovative, and imitative tasks of the entrepreneur. The combination of direct U.S. manufacturing investment in Europe, and the technology transfer activities of multinational corporations in the postwar era were also of great significance (see, for example, David Teece, 1976). All of this, incidentally, suggests that as the forces making for convergence were stronger in the postwar era than previously, the rate of convergence should have been higher. The evidence assembled by Abramovitz (1985) on the basis of Maddison's data indicates that this is in fact what has happened.

The process that has just been described, then, provides mutual benefits, but it inherently helps productivity laggards more than leaders. For the laggards have more to learn from the leaders, and that is why the process makes for convergence.

Like innovation, investment, generally considered the second main source of growth in labor productivity, may also exhibit international public good properties. Suppose two industrialized countries, A and B, each produce two traded products: say automobiles and shoes, with the former more capital intensive. If A's investment rate is greater than B's then, with time, A's output mix will shift toward the cars while B's will move toward shoes. The increased demand for auto workers in A will raise their real wages, while A's increased demand for imports of B's shoes will raise real wages in B, and will raise the *value* of gross domestic product per labor hour in that country. Thus, even investment in country A automatically tends to have a spillover effect on value productivity and real wages *in those other countries that produce and trade in a similar array of goods.*

While, strictly speaking, the factor-price equalization theorem is not applicable to my discussion because it assumes, among other things, that technology is identical in all the countries involved, it does suggest why (for the reasons just discussed) a high investment rate may fail to bring a relative wage advantage to the investing country. In practice, the conditions of the theorem are not satisfied precisely, so countries in which investment rates are relatively high do seem to obtain increased relative real wages.[6] Yet the analysis suggests that the absolute benefits are contagious—that one country's successful investment policy will also raise productivity and living standards in other industrialized countries.[7]

---

[6] The case of Germany and Britain is a suggestive illustration. From 1870 to 1979, according to Maddison (pp. 227, 231) German net nonresidential fixed tangible capital stock grew thirtyfold while that of the United Kingdom grew only sixfold. Landes (p. 124) reports (with the appropriate warning, "*caveat lector*") that real U.K. wages in about 1860 have been estimated to be some 2.5 times those in Germany, while according to the U.S. Bureau of Labor Statistics (unpublished figures), in 1980 German real wages were 1.7 times those in the United Kingdom. Of course, I am not suggesting that correlation implies causation, or that one such case even provides a usable correlation. Still, the orders of magnitude of the figures are probably right and it is hard to believe that superior German investment rates had nothing to do with the relative lag in British wages.

[7] It must be conceded that the longer-run data do not seem to offer impressive support for the hypothesis that the forces of factor-price equalization have, albeit imperfectly, extended the benefits of exceptional rates of investment from those economies that carried out the successful investment programs to other industrialized economies. Since we have estimates of relative real wages, capital stock, and other pertinent variables for the United Kingdom and Germany, these have been compared in Table 2. If the public goods attribute hypothesis about the effects of investment in one country were valid and if factor-price equalization were an effective force, we would expect the relative rise in

Thus, effective growth policy does contribute to a nation's living standards, but it may also help other industrialized countries and to almost the same degree; meaning that relative deviations from the patterns indicated in Figure 2 will be fairly small, just as the diagram shows. (However, see Abramovitz, 1985, for a discussion of the counterhypothesis, that growth of a leader creates "backwash" effects inhibiting growth of the followers.)

All this raises an obvious policy issue. If productivity growth does indeed have such public good properties, what will induce each country to invest the socially optimal effort and other resources in productivity growth, when it can instead hope to be a free rider? In part, the answer is that in Western capitalistic economies, investment is decentralized and individual firms can gain little by free riding on the actions of investors in other economies, so that the problem does not appear to be a serious one at the national policy level.

IV. Is Convergence Ubiquitous?

Does convergence of productivity levels extend beyond the free-market industrialized countries? Or is the convergence "club" a very exclusive organization? While century-long data are not available for any large number of countries, Summers and Heston provide pertinent figures for the 30-year period 1950–80 (data for more countries are available for briefer periods).[8] Instead of labor productivity figures, they give output per capita, whose trends can with considerable reservations be used as a rough proxy for those in productivity, as Maddison's figures confirm.

Figure 3 tells the story. Constructed just like Figure 2, it plots the 1950–80 real growth rates of *GDP* per capita for all 72 Summers-Heston countries against the initial (1950) level of this variable. The points form no tight relationship, and unlike those for the industrial countries, the dots show no negatively sloping pattern. Indeed, a regression yields a slightly positive slope. Thus, rather than sharing in convergence, some of the poorest countries have also been growing most slowly.

Figure 3 brings out the patterns more clearly by surrounding the set of points representing Maddison's 16 countries with a thin boundary and the centrally planned economy points[9] with a heavier boundary. We see that the Maddison country points lie near a sort of upper-right-hand boundary, meaning that most of them had the high incomes in 1950 (as was to be expected) and, for any given per capita income, the highest growth rates between 1950 and 1980. This region is very long, narrow, and negatively sloped, with the absolute slope declining toward the right. As in the Figure 2 productivity data for a 110-year period, this is exactly the shape one expects with convergence. Second, we see that the centrally planned economies are members of a convergence club of their own, forming a negatively sloping region lying below and to the left of the Maddison countries. The relationship is less tight, so convergence within the group is less pronounced, but it is clearly there.

German real wages and in productivity to be small (on some criterion) in comparison with the relative increase in its capital stock. However, the figures do not seem to exhibit such a pattern.

[8] There are at least two sources of such data: the World Bank and the University of Pennsylvania group. Here I report only data drawn from the latter, since their international comparisons have been carried out with unique sophistication and insight. Instead of translating the different currencies into one another using inadequate exchange rate comparisons, they use carefully constructed indices of relative purchasing power. I have also replicated my calculations using World Bank data and obtained exactly the same qualitative results.

[9] The centrally planned economies are Bulgaria, China, Czechoslovakia, East Germany, Hungary, Poland, Romania, USSR, and Yugoslavia. The 5 countries with relatively high 1950 incomes included neither in Maddison's sample nor in the planned group are, in descending order of *GDP* per capita, Luxemburg, New Zealand, Iceland, Venezuela, and Argentina. The countries with negative growth rates are Uganda and Nigeria.

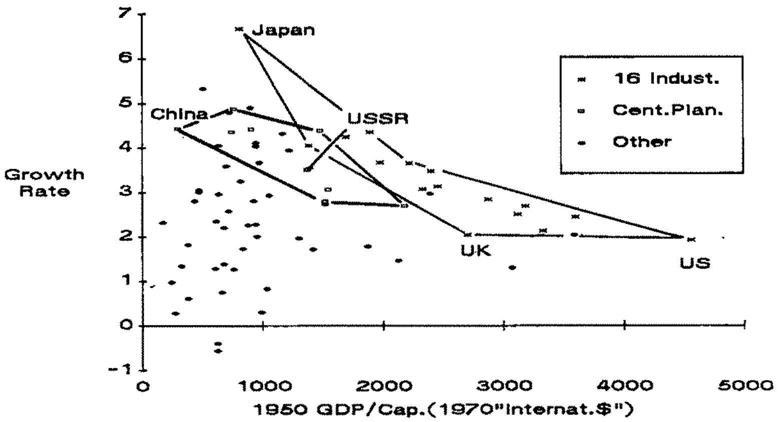


FIGURE 3. GROWTH RATE, 1950–80, GROSS DOMESTIC PRODUCT PER CAPITA VS. 1950 LEVEL, 72 COUNTRIES

Source: Summers and Heston

Finally, there is the region of remaining points (aside from the rightmost non-Maddison points in the graph) which lies close to the origin of the graph and occupies something like a distorted circle without any apparent slope. The points closest to the origin are less developed countries which were poor in 1950, and have grown relatively slowly since. They show no convergence among themselves, much less with other groups.

A few numbers suggest the difference in performance of various subgroups of the 72 countries. Using a four-set classification Summers, I. B. Kravis and Heston (1984, p. 254) provide Gini coefficients by decade from 1950 to 1980. For their set of industrialized countries, this coefficient falls precipitously from 0.302 in 1950 to 0.129 in 1980—a sharp drop in inequality. For the centrally planned economies the drop is much smaller—from 0.381 to 0.301. The middle-income group exhibits an even smaller decline, from 0.269 to 0.258. But the low-income countries underwent a small *rise* over the period, from 0.103 to 0.112, and the world as a whole experienced a tiny rise from 0.493 to 0.498.

There has also been little convergence among the groups. For the entire period, Summers et al. report (p. 245) an average annual growth rate in per capita real *GDP* of 3.1 percent for industrialized countries, 3.6 percent for centrally planned economies, 3.0 percent for middle-income market economies, and only 1.5 percent for the low-income group, with a world average growth rate of 2.7 percent.

This suggests that there is more than one convergence club. Rather, there are perhaps three, with the centrally planned and the intermediate groups somewhat inferior in performance to that of the free-market industrialized countries. It is also clear that the poorer less developed countries are still largely barred from the homogenization processes. Since any search for "the causes" of a complex economic phenomenon of reality is likely to prove fruitless, no attempt will be made here to explain systematically why poorer less developed countries have benefited to a relatively small degree from the public good properties of the innovations and investments of other nations. But part of the explanation may well be related to product mix and education. A less developed country that produces no cars cannot benefit from the invention and adoption of a better car-producing robot in Japan (though it does benefit to a lesser degree from new textile and rice-growing technology), nor can it benefit from the factor-price equalization effect of the accompanying Japanese investments, since it cannot shift labor force out of its (nonexistent) auto industry as the theorem's logic requires. Lack of education and the associated skills prevent both the presence of high-tech industries and the effective imitation (adoption) of the Japanese innovation. Obviously, there is much more to any reasonably fuller explanation of the exclusion of many less developed countries from the convergence process, but that is not my purpose here.

V. The Record of the United States

The long-run data call for a revaluation of the past productivity performance of the United States, which is rather different from what is widely believed. Figure 4 plots for 4 countries growth rates of *GDP* per work-hour derived from Maddison. It confirms, of course, that U.S. labor productivity growth has been lower than that of several other countries in recent decades. But it also indicates that this U.S. growth rate has just been middling for the better part of a century, not only in the past 15 or 20 years. This is confirmed by M-F-O (p. 31) whose data show that between 1899 and 1913, in terms of *GDP* per work year, the U.S. growth rate was already lower than that of Sweden, France, Germany, Italy, and Japan. Its growth rate was also below theirs (except for France) in 1924–37. While U.S. labor productivity grew rapidly relative to other nations during both world wars, this is attributable in good part to a slowdown in the growth of other countries. Thus, the mediocre U.S. relative performance in growth in labor productivity is an old story, not just a postwar phenomenon.

Figure 2—the graph with the inverse correlation between initial productivity level and its subsequent growth in Maddison's countries— suggests an explanation of the relatively undistinguished U.S. performance. The convergence of productivity levels in industrialized countries inevitably condemned those with high 1870 productivity levels to relatively slow growth since then. This attributes much of the modesty of the U.S. productivity growth to the high level it had achieved earlier. Indeed, U.S. performance was notably better than this view suggests. While in 1870, U.S. *GDP* per work-hour ranked fifth from the top among Maddison's 16 countries, its subsequent productivity growth was seventh, not fifth, from the bottom. And while my regression equation between initial productivity and its subsequent growth predicts that the United States should have achieved about a tenfold growth in its labor productivity since 1870, its actual achievement was some 20 percent higher than this. On this interpretation, then, rather than a manifestation of failure, the growth rate of the United States over the course of the 110 years represents a mild achievement in comparison with what might have been expected from the convergence relationship.

Next let us consider the slowdown hypothesis—the assertion that U.S. productivity has fallen sharply below its past. The U.S. figures provided by John Kendrick (see U.S. Bureau of Census, 1973), like the Maddison data, exhibit no major break in the *long-run* U.S. trend and no sign of a long-run slowdown (similar conclusions were reached by Michael Darby, 1984). In Figure 1, the time-series (semilog) representation of Maddison's figures on levels of labor productivity, the curve for the United States is remarkably close to a straight line, except for the dip in the 1930's. A graph of Kendrick's data is almost identical in pattern. In Figure 4, a century of U.S. growth rate in labor productivity clearly exhibits neither a downward long-term trend nor a marked recent dip below its historical level. The mild shortfall shown for the past few years must also be revaluated because the last year in the graph was a period of recession, which usually depresses productivity. More recent data may well wipe out much of the small fall below historical levels.

True, there has undoubtedly been a protracted fall off from the early postwar peak, and it certainly was pronounced. But it is that peak which looks like the aberration, and the decline from it may well prove to be a return to historical growth rates in labor productivity. In this connection, it is noteworthy (Figure 4) that the duration and amplitude of the great leap above historical

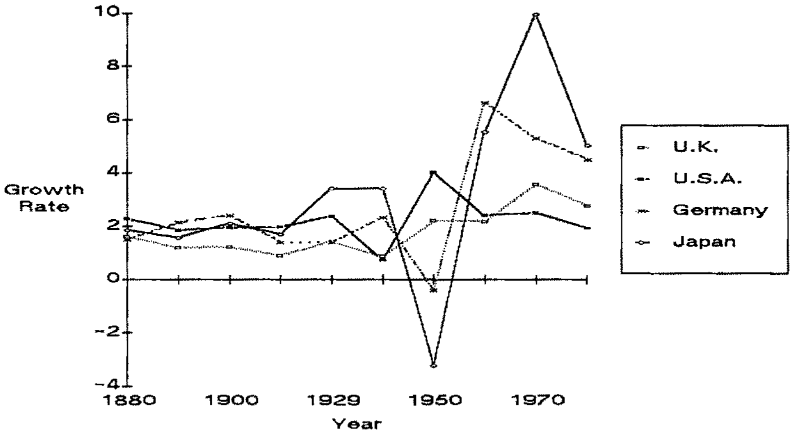


FIGURE 4. GROWTH RATES, GROSS DOMESTIC PRODUCT PER WORK-HOUR, 1870–1979

Source: Maddison (p. 212)

U.S. productivity growth in the war and early postwar years were just about as great as the previous shortfalls during the Great Depression. This encourages reinterpretation of the postwar growth period as one of temporary catch-up, merely making up for opportunities previously foregone. Perhaps the accumulated innovative ideas, unused because of the depression, as well as frustrated savings goals, fueled an outburst of innovation and investment when business conditions permitted. With time, as the backlog of ideas and investable funds was depleted, productivity growth rates declined to their normal levels. (On all this, see Abramovitz, 1979.)

VI. Productivity, Unemployment, and Deindustrialization

The long-term data also permit us to dispose of some other popular views, several of which economic theory has long ago shown to be myths. It is widely feared that rapid labor productivity growth will destroy jobs, even in the long run. Second, somewhat inconsistently, it is feared that if an economy's productivity growth lags behind that of other countries, it will lose jobs to foreign workers, its industry will suffer, and its balance of payments will face chronic deficits (these last alleged consequences are sometimes called "deindustrialization"). Economists will not be surprised that the data do not support any of these conclusions.[10]

The long-run unemployment allegation implies that the twelvefold increase in U.S. output per work-hour and the sevenfold and sixteenfold increases in the United Kingdom and Germany, respectively, must have had a devastating effect on their labor demands. Even with a 50 percent cut in annual work-hours, employment could have fallen by 5/6, relative to population, without a decline in gross domestic product per capita.

The data confirm that nothing of the sort actually occurred. M-F-O's figures for unemployment rates for the United Kingdom, the United States, and Germany from 1874 to 1973 indicate that, while unemployment rates averaged about 4 percent before World War II, they had, if anything, fallen somewhat below that level in the period 1952–73. This is so despite the substantial rise in the ratio of number of persons in the labor force to total population which, according to Maddison's figures, went up in all 3 countries from an unweighted mean of 34 percent in 1870 to 45 percent in 1979.

---

[10] For a systematic study that rejects the deindustrialization theses for postwar United States, see Robert Lawrence (1984).

The same data also undermine the (nearly) opposite apprehension—that a laggard in productivity growth is subject to extraordinary unemployment problems. The M-F-O figures (p. 94) show a fall in U.K. unemployment rates from 4.7 percent from 1873 to 1913 to some 3 percent, on average, in 1952–73. Of course, unemployment rates have since risen sharply, but this is not evidence of a long-term trend.

There is no more substance to the view of many noneconomists that a persistent lag in a nation's productivity growth will place it at a competitive disadvantage in international trade, excluding it increasingly from export markets, with devastating effects upon its export industries and its balance of payments. Economic analysis denies much of this. Balance of payments will ultimately be brought into equilibrium whatever a nation's productivity performance and, since trade depends heavily on comparative rather than absolute advantage, it need have little effect on exports. Instead, the exchange rate *and the standard of living* of the country with lagging productivity will bear the brunt of the burden as it is forced, increasingly, to compete by means of relatively low wages. It is true that the United Kingdom's share of world manufacturing exports has fallen sharply from 43 percent of the world's total in 1880 all the way down to 9 percent of the total in 1973 (M-F-O, p. 425). But Britain's net exports of goods and services (the balance of visible trade) has moved fairly steadily in its favor since the 1870's (M-F-O, p. 443).

In addition, the absolute volume of U.K. exports rose spectacularly. As Table 1 reports, between 1870 and 1979 British exports increased about tenfold, and from 1855 to 1973 U.K. exports of *goods* increased 13 times (M-F-O, p. 427).

Nor has the United Kingdom been forced to deindustrialize internally. Maddison's figures (p. 205) on the share of the labor force in industry do show that Britain declined from first place among his 16 countries in 1870 to fourth place (behind Germany, Austria, and Switzerland) in 1979. But industry's share of employment in the United Kingdom was still 88 percent of that of the leader's (Germany), and the United Kingdom continued ahead of Sweden, France, the United States, Belgium, and Japan.[11] If this is deindustrialization, it certainly is not extreme.

There remains the question whether the U.K.'s decreased share of world exports damaged its economic welfare severely. That is, if Germany, Japan and Italy had not outpaced U.K. productivity growth and so had not increased their share of world trade at the U.K.'s expense, would an average Englishman have been far better off? The easy conclusion that he would is disputed by D. N. McCloskey (at least for the period 1870–1913) in a thoughtful paper (1981, pp. 173–83). First, whatever portion of Britain's output was not exported as a result must have added to domestic consumption or investment. The *net* loss on this score (the net producer's surplus on the exports) may have been negligible. In addition, growing productivity in Japan, Germany and elsewhere enables a British subject to import cameras, TV sets, and many other items far more cheaply than if the productivity of other countries had stagnated. There must also have been an upward shift in the demand curves for British goods by those countries because of their rising incomes. McCloskey concludes that reports of the detrimental effect on the U.K. living standard of the productivity record of its industrialized rivals have been greatly exaggerated.

VII. Concluding Comment

This paper is an exercise in interpretation of data supplied by others. I have examined no primary sources and have not attempted to supply, or even to revise, sets of statistics (though I have undertaken to cross check my results where several parallel data sets were available). Yet the data were able to provide some suggestive conclusions which need not be repeated here.

---

[11] Here, too, there has been a considerable narrowing of range. In 1870, with data for 4 countries missing, the share of workers in industry extended from 9.7 percent (Finland) to 42.3 (United Kingdom), with several countries near 20 percent and several in the high 30's. By 1979, that range ran only from a low of 28.7 percent (Canada) to a high of 44 percent (Germany).

Above all, it seems to me the paper has reaffirmed that the study of economic history is not simply a manifestation of "idle curiosity" (to use Veblen's characterization of the motive for academic research). The long run does matter. It matters in several ways, at least one of them confirmed by the preceding discussion.

First, important current issues are, I believe, the product of path-dependent processes whose mathematical expression must take the form of functionals rather than mere functions, meaning that we cannot understand current phenomena such as the relative productive capacities of different economies without systematic examination of earlier events which affect the present and will continue to exercise profound effects tomorrow. (This, in my view of the matter, is a major element in the Hegelian and Marxian view of the importance of history.)

Second, the long run matters, because policies designed with only short-run problems and consequences in mind are all too likely to backfire once the immediate crisis is past. (This and the next point were, I believe, the main sources of Viner's preoccupation with the long run.)

Third, focus upon short-run phenomena such as recessions may lead an investigator to ignore more powerful and persistent forces such as those that were the primary concern of the classical economists.

Finally, and most pertinent to the discussion here, the long run is important because it is not sensible for economists and policymakers to attempt to discern long-run trends and their outcomes from the flow of short-run developments, which may be dominated by transient conditions. The validity of this view is surely confirmed at several points in this article, perhaps most dramatically by the way history places in perspective the recent developments in U.S. productivity which have been the focus of so much alarm.

REFERENCES

Abramovitz, Moses, "Rapid Growth Potential and its Realization: The Experience of the Capitalist Economies in the Postwar Period," in Edmond Malinvaud, ed., *Economic Growth and Resources, Proceedings of the Fifth World Congress of the International Economic Association*, Vol. I, London: Macmillan, 1979.

______, "Catching Up and Falling Behind," delivered at the Economic History Association, September 20, 1985.

Ames, Edward and Rosenberg, Nathan, "Changing Technological Leadership and Industrial Growth," *Economic Journal*, March 1963, *73*, 13–31.

Beniger, James R., *The Control Revolution: Technological and Economic Origins of the Information Society*, Cambridge: Harvard University Press, forthcoming.

Caves, Richard E. and Krause, Lawrence B., *The Australian Economy: A View From the North*, Washington: The Brookings Institution, 1984.

Cipolla, Carlo M., *Before the Industrial Revolution: European Society and Economy, 1000–1700*, New York: W. W. Norton, 1976.

Clark, Colin, *The Conditions of Economic Progress*, 3rd ed., London: Macmillian, 1957.

Darby, Michael, "The U.S. Productivity Slowdown: A Case of Statistical Myopia," *American Economic Review*, June 1984, *74*, 301–322.

David, Paul A., "Invention and Accumulation in America's Economic Growth: A Nineteenth-Century Parable," in K. Brunner and A. H. Meltzer, eds., *International Organization, National Policies and Economic Development*, Amsterdam: North-Holland, 1977, 179–228.

Deane, Phyllis and Cole, W. A., *British Economic Growth 1688–1959*, Cambridge: Cambridge University Press, 1962.

Feinstein, Charles, *National Income, Expenditure and Output of the United Kingdom, 1855–1965*, Cambridge: Cambridge University Press, 1972.

Gerschenkron, Alexander, "Economic Backwardness in Historical Perspective," in Bert F. Hoselitz, ed., *The Progress of Underdeveloped Areas*, Chicago: University of Chicago Press, 1952.

Landes, David S., *The Unbound Prometheus*, Cambridge: Cambridge University Press, 1969.

Lawrence, Robert Z., *Can America Compete?*, Washington: The Brookings Institution, 1984.

McCloskey, D. N., *Enterprise and Trade in Victorian Britain*, London: Allen & Unwin, 1981.

Maddison, Angus, *Phases of Capitalist Development*, New York: Oxford University Press, 1982.

Marx, Karl and Engels, Friedrich, *Manifesto of the Communist Party* (1848), London: Lawrence and Wishart, 1946.

Matthews, R. C. O., *Slower Growth in the Western World*, London: Heinemann, 1982.

______, Feinstein, C. H. and Odling-Smee, J. C., *British Economic Growth, 1856–1973*, Stanford: Stanford University Press, 1982.

Mueller, Dennis C., *The Political Economy of Growth*, New Haven: Yale University Press, 1983.

Phelps Brown, E. H. and Hopkins, S. V., "Seven Centuries of Building Wages," *Economica*, August 1955, *22*, 195–206.

______, and ______., "Seven Centuries of the Prices of Consumables," *Economica*, November 1956, *23*, 296–314.

Porat, Marc Uri, "The Information Economy, Definitions and Measurement," Office of Telecommunications, Special Publication 77–12(1), U.S. Department of Commerce, Washington, 1977.

Romer, Paul M., "Increasing Returns and Long Run Growth," Working Paper No. 27, University of Rochester, October 1985.

Summers, Robert and Heston, Alan, "Improved International Comparisons of Real Product and its Composition, 1950–1980," *Review of Income and Wealth*, June 1984, *30*, 207–262.

______, Kravis, I. B and Heston, Alan, "Changes in World Income Distribution," *Journal of Policy Modeling*, May 1986, *6*, 237–269.

Teece, David J., *The Multinational Corporation and the Resources Cost of International Technology Transfer*, Cambridge: Ballinger, 1976.

Veblen, Thorstein, *Imperial Germany and the Industrial Revolution*, New York: Macmillan, 1915.

Viner, Jacob, *The Long View and the Short*, Glencoe: Free Press, 1958.

Williamson, Jeffrey G., "Why was British Growth So Slow During the Industrial Revolution?," unpublished, Harvard Institute of Economic Research, 1983.

U.S. Bureau of Census, *Long Term Economic Growth* 1860–1970, Washington, June 1973.

Taxation of Investment and Savings in a World Economy

By ROGER H. GORDON*

The equilibrium allocation of capital and equilibrium market prices are derived for a world economy with a unified securities market, mobile capital, no uncertainty, and varying tax rates on different sources of income in each country. The paper then characterizes optimal tax rates for a small country in this setting, focusing on the peculiar incentives created when the before-tax rate of return differs among securities due to differences in their typical tax treatment.

Traditionally, papers examining the effects of taxes on the amount of savings, the allocation of capital, the composition of individual portfolios, and economic efficiency, have assumed that the economy is closed. There have been a few recent exceptions to this rule, among them Martin Feldstein and David Hartman (1979), Hartman (1981, 1985), Daniel Frisch (1983), and Lawrence Goulder, John Shoven, and John Whalley (1983). These papers, however, have generally made quite restrictive assumptions about the determinants of international movements of capital. For example, Feldstein and Hartman, Hartman (1981), and Frisch restrict their discussion to the incentives faced by multinationals based in the United States who consider investing abroad, while Hartman (1985) focuses on direct investment in the United States by foreigners. Goulder et al. adopt a more flexible specification in which movement of capital depends on interest rate differentials, but not on taxes.

The first objective of this paper is to develop a general equilibrium model of how taxes affect the amount and form of international flow of capital. Among other questions, it will examine how taxes affect relative interest rates in different countries, the ownership pattern of financial and real assets, and relative capital intensities in different countries. The model also suggests a possible tax explanation for the importance of multinationals. The other objective of the paper is to analyze the implications of this model for the optimal tax policy of a small open economy.

Throughout the paper, in order to keep the model simple, I assume that there is no uncertainty, that all firms have the same technology, and that capital and output are freely mobile between countries. To obtain an equilibrium, I assume a sufficient set of constraints on portfolio behavior, following the approach used in Merton Miller (1977). As in Miller, each individual will either be completely indifferent or prefer a corner solution when choosing between any pair of assets.[1]

In Section I, I develop a model of equilibrium savings, portfolio composition, and capital allocation in an open economy assuming either that all assets are traded internationally, or, alternatively, that only bonds are traded internationally. In Section II, I explore the implications of each of these models for optimal tax policy. Finally, Section III provides a summary of the results and a short discussion of the effects of relaxing some of the less realistic assumptions.

---

*Professor of Economics, University of Michigan, Ann Arbor, MI 48109 and Research Associate, NBER. I thank Mary Lovely, Jack Mintz, Joel Slemrod, Michelle White, the referees, participants of seminars at Illinois, Indiana, Michigan, NBER, and Toronto, and especially Jay Wilson for comments on an earlier draft. I also thank Hal Varian for assistance with the graphics. Much of this paper was written while I was visiting Tel-Aviv University and the Sonderforschungsbereich 303 at the University of Bonn.

[1] Harvey Galper and Eric Toder (1984) used a similar set of assumptions to analyze savings and investment incentives in a closed economy with many assets, each treated differently for tax purposes.

I. Equilibrium Portfolios and Capital Allocation

A. *Characteristics of the Model*

In this world economy, there are N countries. Each country consists of one representative individual and a corporate sector which acts competitively. The economy operates for two periods. There is only one good in the world economy, which is freely tradeable among the countries and can be either consumed or invested. This assumption implies that there will be purchasing power parity in each period. By assumption, the price level is the same in the first period in every country, and is the numeraire.

In the first period in each country j, the government produces an amount of public services, G_j, financed by debt, which it repays out of tax revenues in the second period. The representative individual in country j starts off the first period with wealth Y_j, which can be spent on first-period consumption, C_{j1}, or invested in whatever asset yields the best net return. In addition to government bonds, the individual can invest in corporate bonds or corporate equity. The corporations issue these securities during the first period to finance their capital investments.

During the second period, the individual works L_j hours, earning a before-tax real wage rate of w_j. The tax rate on labor income is s_j. The individual's consumption during the second period, C_{j2}, equals his(her) net-of-tax labor income plus the value of his savings. If the best available real net-of-tax return available on investments in the first period is q_j, then the individual's budget constraint is $C_{j2} = w_j(1-s_j)L_j + (Y_j - C_{j1})(1 + q_j)$. The individual chooses C_{j1} and L_j, subject to this budget constraint, to maximize $W_j = U(C_{j1}, C_{j2}, L_j) + U^*(G_j)$. His resulting utility can be represented, using the indirect utility function, by $V(w_j(1-s_j), q_j; Y_j, G_j)$.

TABLE 1—DEFINITIONS OF SYMBOLS

x_j	Value of some Variable "x" in Country j
C_n	Consumption in Period n
G	Government Expenditures in Period
K	Capital Investment Used in Production in Period 2
L	Labor Supply in Period 2
N	Number of Countries
Y	Initial Wealth
d	Fraction of Capital Financed with Bonds
d_x	Maximum Debt-Capital Ratio
e	Nominal Capital Gains on Equity
g	Tax Rate on Capital Gains
p	Before-Tax Return on Individual's Investments
p^c	Before-Tax Cost of Finance of Corporate Investments
q	Best Available Net-of-Tax Return to Investors
q^c	Minimum Net-of-Tax Cost of Finance for a Corporation
r	Nominal Interest Rate
s	Tax Rate on Labor Income
t	Tax Rate on Interest Income
w	Before-Tax Wage Rate
γ	Tax Rate on Corporate Capital Gains
π	Domestic Inflation Rate
τ	Tax Rate on Corporate Income
i	$(r-\pi)/(1+\pi)$
v	$1/(1+\pi)$
α	$(1-t)/(1-g)$
α^c	$(1-\tau)/(1-\gamma)$

When investing in the first period, each individual can invest in government bonds, corporate bonds, or corporate equity issued in any of the N countries. Bonds issued in any country k pay a nominal coupon rate denoted by r_k. This nominal coupon is taxable in the investor's home country j at a tax rate t_j, where $t_j > 0$.[2] Let the inflation rate in country k between the two periods be denoted by π_k. As long as the inflation rates in countries j and k differ, then the investor will have a capital gain or loss due to exchange rate changes when the bond principal is repaid. This capital gain or loss will be subject to tax at the capital gains tax rate equal to g_j, where $g_j < t_j$. Given this assumption, a unit invested in bonds of country k has a real net-of-tax value in the second period of $(1+r_k(1-t_j))/(1+\pi_k) - g_j[1/(1+\pi_k) - 1/(1+\pi_j)]$. Let v_k represent the real value of a unit of currency in country k in the second period for each k, so that $v_k = 1/(1+\pi_k)$. Then the real net-of-tax value of a unit invested in bonds can be reexpressed as $1+v_k[r_k(1-t_j)-\pi_k(1-g_j)]-v_jg_j\pi_j$.

---

[2] I ignore here any withholding tax due in the country in which the bonds were issued. For a discussion of the use of withholding taxes, see Vito Tanzi (1984, pp. 62–63).

If the individual from country j invests instead in equity issued in country k, the entire return he receives is assumed to take the form of capital gains, which are taxable at the same capital gains tax rate g_j.[3] Let e_k denote the nominal rate of capital gains on equity issued in country k. Then the real after-tax value in the second period of a unit invested in this security equals $v_k(1+e_k)-g_j[v_k(1+e_k)-v_j]=1+v_k(e_k-\pi_k)(1-g_j)-v_jg_j\pi_j$.

In order to have a well-defined equilibrium in the securities market in this setting, some constraints must be imposed on individual trading.[4] Following Miller, I assume that individuals must invest a nonnegative amount in an asset, so that by assumption they cannot borrow and cannot sell equity short. Some justification for these assumptions can be found in institutional limitations in the equity market, and in tax law regulations which at least in the United States prevent unlimited interest deductions and rule out capital gains treatment of short sales of equity. Each individual would then choose to invest in that security or securities yielding the highest available after-tax return.

In each country j, corporations issue equity and bonds during the first period to finance their capital stock. These securities can be issued in any currency. Let d_j represent the fraction of the capital financed with bonds. I assume that this fraction d_j must be nonnegative and must not exceed some value d_x. These assumptions are necessary to guarantee an equilibrium. U.S. tax law prevents a corporation from investing in more than a small amount of financial assets, preventing d from becoming too negative. Also, agency and bankruptcy costs presumably prevent corporations from borrowing too heavily. Implicitly, I am assuming that each firm has idiosyncratic risk which limits its ability to borrow but does not give rise to any market risk.

During the second period, the corporation uses this capital along with labor to produce output using a constant returns to scale production function. If the amount of capital raised is K_j, then real output, net of depreciation, is $f(K_j, L_j)$, where $f_K > 0$ and $f_{KK} < 0$. This output, net of labor expenses, is subject to a corporate income tax at rate τ_j.[5] This leaves a net income of $(f(K_j, L_j)-w_jL_j)(1-\tau_j)$, plus the original capital stock K_j, which together is used to pay back the firm's equity holders and bondholders.[6]

Coupon payments to bondholders are assumed to be deductible from corporate taxable income. In addition, any capital gains or losses on the amount that must be repaid as principal on these bonds is subject to taxation at a corporate capital gains tax rate of γ_j. If the firm originally borrowed an amount d_jK_j in some country k, then the amount of real goods it must have available in the second period in order to repay this debt, given the tax treatment of coupon payments and capital gains, equals $d_jK_j[1+v_k(r_k(1-\tau_j)-\pi_k(1-\gamma_j))-v_j\gamma_j\pi_j]$.[7]

Similarly, if the firm raised the remaining amount of capital, $(1-d_j)K_j$, by issuing equity in some country k, then the amount of real goods it needs in the second period to provide the rate of return required in the market equals $(1-d_j)K_j[1+v_k(e_k-\pi_k)]$. The amount it pays to equity holders and to bondholders together must equal the amount of resources it has available.

---

[3] I assume for simplicity that the tax rate on exchange rate capital gains or losses is the same as that on other types of capital gains or losses. Some countries do vary the tax rate on capital gains, depending on the source.

[4] For a discussion of the necessity of constraints on behavior in a certainty model with different tax rates, if infinite amounts of tax arbitrage are to be avoided, see Mervyn King (1977).

[5] To the degree that the actual corporate tax base differs from the one assumed here, such as by having historic cost depreciation or investment credits, then the tax rate on f_K would have to be modified appropriately. However, the statutory rates would still apply to interest payments and exchange rate gains or losses.

[6] I ignore here any tax complications that might arise if the corporation is a multinational based elsewhere. In principle, the income net of local taxes might be subject to further taxes when it is repatriated to the home country. As Hartman (1981) argues, a U.S.-based corporation can avoid these further taxes either by indefinitely postponing repatriation, or by investing in such a mix of foreign countries that its average tax payment abroad is at least as high as the U.S. tax payment would be. These complications are discussed further below.

[7] This follows from the same derivation used in calculating the real amount an individual investor receives from owning a bond.

I assume that each corporation is a price taker in the world securities market, and issues that type of security and in that country which minimizes its cost of funds, subject to the constraints on d_j described above. Let $1+q_j^c$ represent the minimum real amount the firm in country j must pay back in the second period per unit of funds raised in the first period. The firm is assumed to choose the amount of funds, K_j, that it raises in the first period so as to maximize profits, implying that in equilibrium $f_K(1-\tau_j)=q_j^c$.

B. *Equilibrium in a Closed Economy*

Consider first the equilibrium of this model in the special case in which there is only one country. This case describes a closed economy, the context of most of the papers analyzing the effects of taxes on savings and investment. Given the necessary presence of government bonds and corporate equity, equilibrium in the securities market in country j requires that the investor in that country be willing to own both these domestic securities. Therefore, the after-tax rate of return on these two securities must be the same, implying that $(1-t_j)r_j=(1-g_j)e_j$ or, equivalently, $\alpha_j r_j=e_j$, where by definition $\alpha_j=(1-t_j)/(1-g_j)$. Here α_j captures the tax disadvantage of coupon payments, which are taxed as ordinary income, relative to other income which takes the form of capital gains.

Given this relation, it is trivial to show that the corporation will find debt finance cheaper than equity finance if and only if $\tau_j>1-\alpha_j$ or, alternatively, if and only if $\tau_j+g_j(1-\tau_j)>t_j$.[8] Hence, the left-hand side of the second inequality describes the percent of income paid in taxes under equity finance through both the corporate tax and the capital gains tax applied to that income remaining after corporate taxes, while the right-hand side describes the percent of income paid in taxes under debt finance. If debt is cheaper, the firm will set d_j equal to its maximum possible value d_x; otherwise it will use solely equity finance. With whatever value it chooses for d_j, the optimal capital stock for the firm will be characterized implicitly by[9]

$$(1)\quad f_K(K_j,L_j)(1-\tau_j) = \alpha_j r_j(1-d_j)+d_j r_j(1-\tau_j)-\pi_j.$$

C. *Equilibrium in an Open Economy*

Assume now that there are N countries in the world economy, with individual investors in each country able to invest in any of $2N$ assets: bonds and corporate equity issued in each of the N countries. For convenience, let these countries be indexed based on their inflation rates; those that have low inflation rates will have a low index. This implies that $v_1>v_2>\ldots>v_N$, where $v_j=1/(1+\pi_j)$.

In equilibrium, all assets must be owned, each investor buys only those assets which yield the highest net return, and each corporation finances itself in the cheapest available way. An investor in country j must compare the net return on equity issued in each country k, which yields $v_k(e_k-\pi_k)(1-g_j)-v_j g_j\pi_j$, and the net return on bonds in each country k, which yields $v_k[r_k(1-t_j)-\pi_k(1-g_j)]-v_j g_j\pi_j$. A number can be added to each of these yields, and each of these yields can be multiplied by some (positive) number, without changing the ranking of these $2N$ available returns. Therefore, an equivalent statement is that the investor is looking for the maximum of the $v_k(e_k-\pi_k)$'s and the $v_k(\alpha_j r_k-\pi_k)$'s. Since coupon payments are taxed more heavily, they must be discounted—receiving a unit of coupon payments is equivalent to receiving α_j units of capital gains. All other income is already taxed as a capital gain.

---

[8] This is a standard result in corporate finance. See, for example, my earlier paper (1982).

[9] Similar results appear in Alan Auerbach (1979) or my article (1984).

It immediately follows that all equity that is issued must earn the same real before-tax return, that is, all the values of $v_k(e_k - \pi_k)$ must be the same. If some value were lower, then all investors would find this security dominated by others that are available. It also follows easily that the real before-tax return earned on each bond must exceed the common real return earned on equity in all countries, so that $v_k(r_k - \pi_k) > v_k(e_k - \pi_k)$. To see this, assume to the contrary that this inequality does not hold for bonds from some country k. Then, since $0 < \alpha_j < 1$ for all j, every investor would find bonds from this country k dominated by equity and the market for this bond would not clear. Investors need some compensation when they buy a bond because some of the income earned on that bond is taxable at the ordinary tax rate, t_j, rather than at the capital gains tax rate, g_j.

By the same reasoning, it follows that bonds issued in countries with a higher inflation rate must pay a higher before-tax real interest rate. These bonds must pay a higher nominal coupon to compensate for the greater capital loss on the bond principal caused by the higher inflation rate. But this means that these bonds generate more income taxable at ordinary tax rates, and so must pay a higher real return before tax to compensate for the higher taxes due. To show this formally, consider bonds from two countries, j and k, where the inflation rate is higher in country k. In equilibrium, someone must choose to own the bonds issued in country k, so that for investors from some country m it must be the case that $v_k(\alpha_m r_k - \pi_k) \geq v_j(\alpha_m r_j - \pi_j)$. Therefore, $v_k(r_k - \pi_k) \geq v_j(r_j - \pi_j) + (1 - \alpha_m)(v_k \pi_k - v_j \pi_j)/\alpha_m$. Since the inflation rate is higher in country k than in country j, the last term must be positive, and the result follows.

Several further characteristics of an equilibrium for this economy are described graphically in Figure 1. For simplicity, I deal with only three countries in the graph, the first with a low inflation rate, the second with an intermediate inflation rate, and the third with a high inflation rate. The arguments immediately generalize to many countries. In the graph, the relative returns earned on the available securities are plotted as a function of the tax parameter α faced by an investor.

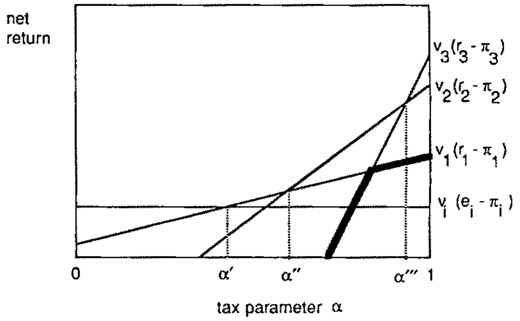


Figure 1. Determination of Investor's Portfolio Choice

Each investor compares the values of $v_k(e_k - \pi_k)$ and $v_k(\alpha r_k - \pi_k)$ for all k when deciding how to invest. The return on equity from each country k must be the same, and does not vary with the tax parameter α. It is therefore plotted in Figure 1 as a horizontal line. The net return earned on each of the bonds increases linearly in α. Since bonds from a country with a higher inflation rate must pay a higher real return, the net return when $\alpha = 1$ must be higher, and so the intercept of the net return line on the right-hand vertical axis must be higher. In addition, since a bond from a country with a higher inflation rate must pay a higher coupon,[10] and so generates more income taxable at ordinary rates, the net return from this bond is more sensitive to the value of the tax parameter α, and the return line in the graph must have a steeper slope. Therefore, the net return lines must look as they do in Figure 1.

Each investor chooses that asset that yields the highest return, given his value of α. As is seen in the figure, investors with values of α

---

[10] This bond must pay a higher coupon not only to compensate for the greater depreciation of the principal due to inflation, but also to compensate for the larger resulting taxable income.

below some α' earn the highest net return by investing in equity. These are the tax payers facing the largest tax penalty for earning income in the form of coupon payments rather than capital gains. Similarly, those investors with values of α above some α''' earn the highest net return by investing in bonds from the country with the highest inflation rate. These bonds pay the largest pre-tax real return, and if the investor faces relatively equal tax rates on coupon payments and capital gains, they also pay the largest post-tax return. Among the remaining investors, those with a value of α between α' and α'', who face a relatively large tax penalty for earning income in the form of coupon payments, prefer the bonds from the country with the low inflation rate, since these bonds generate relatively less income taxable at ordinary rates. In equilibrium, the intercepts of all the asset lines must adjust so that demand for each asset equals the supply of each asset.

This equilibrium is directly analogous to that commonly described for the municipal bond market, where high tax bracket investors buy tax exempt securities. Here, equity is the most tax-favored security since none of its income is taxed as ordinary income, while the bond from the highest inflation country is the least tax favored since it generates a lot of fully taxable income.

The best available asset for an investor is a function solely of the tax parameter α that the investor faces. Unless a country's tax parameter α and its inflation rate are aligned just right, this best asset will be a foreign security.[11]

One other important observation to make from Figure 1 is that there is no one value of the tax parameter α at which all net returns are equal. Explaining differences in real interest rates is more complicated. For example, the real returns on the bonds from the two countries with the highest inflation rates are equal for an investor facing a tax parameter α''', implying, using the above derivation, that $v_3(r_3-\pi_3)=v_2(r_2-\pi_2)+(1-\alpha''')(v_3\pi_3-v_2\pi_2)/\alpha'''$.[12] However, a different tax parameter would need to be used to explain the difference in interest rates between any other pair of bonds, or between any bond and equity. Interest rate parity conditions in this setting are much more complicated than they are in standard trade models. In particular, real before-tax interest rates are not equated across countries in this setting, nor are real after-tax interest rates for any one given tax rate.[13]

Corporations are simultaneously making an analogous set of choices concerning how to finance their investments. A corporation in each country j would compare the amount it must repay if it raises a unit of funds with equity or with bonds in each country k, and choose the minimum cost source of funds, subject to the constraints on its use of debt. The cost of equity finance in each country k is simply $v_k(e_k-\pi_k)$. But this real rate of return required on equity must be the same in all countries, to insure that investors are willing to hold all the issued equity. Therefore, corporations are also indifferent concerning where they issue their equity. They might as well issue their equity domestically.

The choice concerning where, if at all, to issue debt is much more complicated. If a firm in country j were to issue debt, it would do so wherever debt is cheapest, and so would seek the lowest value of $v_k(r_k(1-\tau_j)-\pi_k(1-\gamma_j))-v_j\gamma_j\pi_j$. Equivalently, by the same argument used in analyzing individual portfolio choices, it would choose that country k where $v_k(\alpha_j^c r_k-\pi_k)$ is smallest, where $\alpha_j^c=(1-\tau_j)/(1-\gamma_j)$. The minimum cost of debt is a function of the tax parameter α^c, and is described in Figure 1 by the lower envelope of the three bond return lines. This lower envelope has been highlighted.

---

[11] Maurice Levi (1977) also pointed out in a two-country setting the possibility that the tax law would lead investors in each country to invest in bonds issued in the other country.

[12] This condition can be rewritten in a more natural form as $v_3(r_3(1-t''')-\pi_3(1-g'''))=v_2(r_2(1-t''')-\pi_2(1-g'''))$, where t''' and g''' are tax rates consistent with α'''.

[13] A number of writers, including Hartman (1979), Tanzi (1980), John Makin (1984), and Robin Boadway, Neil Bruce, and Jack Mintz (1984), have pointed out that after-tax real rates, rather than before-tax rates, ought to be equated in equilibrium. These papers assume, however, that there is some one tax rate at which all after-tax real rates are equated, contrary to the above results.

The figure indicates that a corporation, if it were to borrow at all, would do so either in the highest or in the lowest inflation rate country—regardless of the value of α^c, the minimum of the three bond lines is never on the line representing bonds from the country with an intermediate inflation rate. If a corporation faces a low value of α^c, it faces a substantial tax incentive to make payments in the form of coupon payments rather than capital gains to investors, so would prefer to issue bonds in the highest inflation rate country. Conversely, if it faces a high value of α^c, the form of payment makes little difference for tax purposes, and it would prefer to issue bonds in the lowest inflation rate country, where the before-tax real cost is lowest. This result is a general one. To see this, assume to the contrary that for some α^c, a corporation would find it cheapest to borrow in some other country k. Therefore, $v_k(\alpha^c r_k - \pi_k) < v_1(\alpha^c r_1 - \pi_1)$, where the first country has by construction the lowest inflation rate. Since the real coupon rate must be higher in country k, this inequality will continue to be satisfied for any value of α smaller than α^c. Similarly, by assumption, $v_k(\alpha^c r_k - \pi_k) < v_N(\alpha^c r_N - \pi_N)$, where country N has the highest inflation rate. This inequality must continue to be satisfied for any value of α larger than α^c, implying that bonds from country k would not be purchased by any investor, a contradiction.

This result that a corporation, if it borrows at all, prefers to borrow in the highest or the lowest inflation rate country clearly contradicts the evidence. Presumably the agency costs of debt are less if the debt is used to finance an investment made in the same country, and enough so that this consideration normally outweighs the tax incentives.

In order to decide whether to borrow at all, a corporation in country j would compare the cost of equity finance, $v_j(e_j - \pi_j)$,[14] with the minimum cost of debt finance, $v_k(r_k(1-\tau_j)+\pi_k(1-\gamma_j)) - v_j\gamma_j\pi_j$, where k equals either 1 or N. It would use debt finance only if debt finance is cheaper than equity finance.

Given the equilibrium pricing of securities described in Figure 1, we know that the before-tax cost of debt must be greater than the before-tax cost of equity. Since payments to equity holders are not tax deductible, the after-tax cost of equity equals the before-tax cost. However, a firm pays its bondholders through coupon payments which are fully tax deductible, and exchange rate capital losses on the bond principal, which are deductible at the capital gains tax rate. For each unit it borrows in currency k, it saves in taxes an amount equal to $\tau_j v_k(r_k - \pi_k) + (\tau_j - \gamma_j)v_k\pi_k + \gamma_j v_j \pi_j$. This expression is positive at least as long as the before-tax real interest rate is positive, and is larger in countries with higher values of τ_j and π_j.[15] A firm will find it attractive to borrow if these tax savings are large enough to offset the higher before-tax payment that must be made on debt relative to equity finance.

In equilibrium, the cost of funds q_j^c to a firm located in some country j is a weighted average of the required rate of return on equity and the minimum available required rate of return on debt, with the weights being either all on equity, or only $1-d_x$ on equity and d_x on debt. The firm, acting competitively, would then invest until the after-tax marginal product of capital equals this cost of funds q_j^c. Ignoring the possibility of debt finance, all firms face the same cost of funds, regardless of the domestic tax rates faced by individual investors. Due to debt finance, however, those firms located in countries with higher inflation rates and higher corporate tax rates gain more by having access to the bond market and so face a lower cost of funds.[16] As a result of this variation in q_j^c, firms in countries with a higher tax rate need not necessarily have a higher marginal product of capital.

---

[14] Since the cost of equity finance is the same everywhere, I assume without loss that the firm issues equity domestically.

[15] The effect of variations in γ_j on these tax savings is more complicated, but it is easy to show that the tax savings are largest when $\gamma_j = 0$.

[16] The cost of debt finance in any country k, $v_k(r_k(1 - \tau_j) - \pi_k(1 - \gamma_j)) - v_j\gamma_j\pi_j$, is declining in both τ_j and π_j.

In deriving the required rate of return on investment in a country, I did not consider explicitly the possibility of investment by multinationals. All investment in a country, whether by a local company or by a multinational, is subject to local corporate taxes and no personal taxes. If a multinational must pay any additional corporate taxes to its home country at repatriation, then it would require a higher marginal product of capital than the local corporation would in order to be willing to invest, and so could not compete. Under U.S. law, repatriation cannot result in a drop in tax payments. Therefore, with this model, at best multinational firms are at no competitive disadvantage, and would normally be driven out by local corporations.

D. *Equilibrium in an Open Economy with No International Trade in Equity*

The model described above has many counterfactual implications. First, it implies that investors would normally specialize in a foreign security, and that a firm's equity would normally be owned by foreigners. It also forecasts that multinational firms will not be important, since they are at a competitive disadvantage if any taxes are levied by the home country when profits are repatriated.

Simply adding uncertainty to the model would not eliminate these anomalies. While uncertainty would certainly lead investors to diversify their portfolios, it would result in much more international diversification than is observed empirically.[17] A variety of stories might be told explaining why so little international diversification occurs. For example, Michael Dooley and Peter Isard (1980) describe the effects of capital controls on security prices.[18] Risk of default or expropriation may also limit international diversification. Under each of these stories, ownership of both foreign bonds and foreign equity would be limited. However, while some international trade in equity exists, volume is small relative to the volume of international loans. One story that might be told to explain the degree to which equity is owned domestically is that domestic investors are far better informed about the distribution of possible returns on domestic equity than are foreign investors.[19] If more than a small number of poorly informed foreigners invest in a domestic equity market, then they should earn a lower return than domestic investors.[20]

Given the lack of international diversification observed in equity markets, in this section I explore an alternative model in which individuals can buy equity in only domestic firms, but can buy bonds issued anywhere.[21] Under this alternative approach, there will be an interesting tax explanation for the importance of multinationals,[22] and this model will have quite different implications for optimal tax policy.

To begin with, however, I assume in addition that there are no multinational firms. The equilibrium in this setting is very similar to that described previously. For simplicity, assume that each country's investors have more than enough wealth to buy all the domestic equity. Each investor will buy, in addition to equity, whatever bonds are most appropriate given their tax rate. By the same arguments as before, those with high tax rates will prefer low coupon bonds, and conversely. The previous results, describing relative interest rates, will continue to hold.

---

[17] With uncertainty in the output of each firm, investors would deviate from owning a proportional share in the world's market portfolio due to tax incentives, which push them to concentrate their portfolio in those assets described above which are most appropriate given their tax rate. Domestic securities would normally be a minor part of an investor's portfolio. If real exchange rates were also uncertain, Michael Adler and Bernard Dumas (1983) show that, ignoring taxes, investors will purchase a combination of domestic bonds and the market portfolio of equities, so again will not concentrate in domestic equity.

[18] See below, and my paper with Hal Varian (1986) for reasons why a government may choose to institute such capital controls.

[19] Arminio Fraga (1984) describes a different approach, in which each country produces a different commodity, and residents in each country consume primarily the locally produced commodity. With uncertainty, these residents would tend to hedge against uncertainty in the price of this good by investing heavily in domestic securities.

[20] See Albert Kyle (1984) for an analysis of this type of model.

[21] No attempt is made, however, to build in explicitly any reason why international trade in equity may be limited. Introducing differential information to the model, or both uncertainty and many different commodities as in Fraga, would severely complicate the analysis.

[22] See Adler-Dumas (1975) for a similar story of the role of multinationals in compensating for the lack of particular securities markets.

The allocation of capital will be very different in this setting, however. Equity in a country must provide the same real rate of return as is earned on the bonds most appropriate for domestic investors, given their tax rate. This required rate of return on equity will vary by country, being higher in countries where investors face lower tax rates. Countries with high personal tax rates will therefore be *more* capital intensive than countries with low tax rates. This is true regardless of the impact of a country's high personal tax rate on the domestic savings rate. When investors face high personal tax rates, they value more highly the capital gains tax treatment of investments in corporate capital, and are willing to accept in return a lower before-tax return. This result is contrary to the standard results in a closed-economy setting.

As a result, there would be an important role for multinationals in this model. Firms based in countries where investors face a low tax rate must provide a higher real rate of return on equity than firms based elsewhere. Yet all firms investing in a country, by assumption, have the same technology, and, ignoring any surtax at repatriation, pay the same corporate tax rate. Therefore, firms based in countries with high personal tax rates have a competitive advantage when investing abroad. If there were no tax at repatriation, then the equilibrium would be exactly that described earlier. Investors in the high-tax-rate countries would own the equity of domestic firms which invest throughout the world, while investors in lower-tax-rate countries would own bonds of the appropriate type.[23]

For the same reasons that individual investors may not buy equity abroad, however, multinationals may not have quite the same technological opportunities in foreign countries. Local firms may have much better information about local markets than would a foreign entrant.[24] A model which seriously addresses the implications of differential information should therefore end up with some multinationals, given the tax incentives described above, but multinationals would not expand to the point where they control production in all countries with lower personal tax rates.

II. Implications for Government Policy[25]

A. *Optimal Tax Policy with International Trade in All Securities*

In general, characterizing optimal government policy in an open economy is quite complicated, since government decisions can affect world prices, and also may induce other governments to change their policies in response.[26] The objective of this paper is not to describe all these general equilibrium complications, but merely to point out various considerations that have been ignored in the existing literature. To make these considerations clear, I assume: 1) there are many countries and each country is sufficiently small that its policy has a negligible impact on the level of world interest rates; 2) the set of available bonds is sufficiently dense that it can be well approximated by a continuum;[27] and 3) each government sets its policy based on world prices, and does not react strategically to decisions of other governments.

---

[23] If firms based in a particular country must pay a large enough surtax when profits are repatriated, then these firms will invest only domestically, and individual investors there will end up owning the appropriate bonds, as well as equity.

[24] However, a foreign entrant would likely have much better information about foreign markets than would an individual investor about foreign equity.

[25] I would like to thank Jay Wilson for helpful advice on this section.

[26] For an examination of these complications, see Feldstein and Hartman.

[27] Formally, in order to use differential analysis of tax changes, it must be assumed that the before-tax earnings on the investors' portfolio change continuously, if at all, with respect to changes in tax parameters.

Given these assumptions, each country j takes as given the schedule of interest rates set in the world market, as portrayed in Figure 1. In particular, the country faces a horizontal supply of foreign equity at the real rate of return to equity required in the world securities market. In addition, the government faces a given real interest rate at which it can borrow money. Since this real borrowing cost depends on the domestic inflation rate, I denote it by $i(\pi_j)$, so that $i = v_j(r_j - \pi_j)$. Similarly, domestic investors can invest as much as they want in whatever asset is appropriate, given their tax rate. Denote the real (before-tax) return from the best available asset by $p_j(\alpha_j)$. Also, denote the real (before-tax) cost of finance for domestic corporations, given the means of finance they choose, by $p_j^c(\alpha_j^c, \gamma_j)$.

The government in country j is assumed to choose jointly the tax rates on labor income, corporate income, interest income, and both corporate and personal capital gains, using the tax revenue to fund government expenditures, so as to maximize the utility of the domestic resident. In addition, the government has control over the country's inflation rate. I ignore, though, any revenue collected by the inflation tax and focus only on the implications of the inflation rate for market prices and for the revenue collected by the other taxes. I omit the subscripts j in the following discussion except when they are needed for clarity.

As indicated earlier, the indirect utility of the domestic resident can be expressed as $V(w^n, q; Y, G)$, where w^n is defined to equal the net of tax wage, $w(1-s)$, and where q is the net-of-tax rate of return earned on the best available asset. Similarly, the individual's optimal labor supply, L, and optimal consumption in each period, denoted C_1 and C_2, can also be expressed as functions of the same variables.

In solving for its optimal policies, the government maximizes its residents' utility subject to the aggregate resource constraint of the economy:

$$(2) \quad C_2 + (1+i(\pi))G + (1+p^c)K \leq f(K,L) + K + (Y - C_1)(1+p(t)).$$

By assumption, corporate capital investments are made by foreigners, while domestic residents invest in those foreign assets most appropriate given their tax rate.[28] Total domestic income is used to repay foreign owners of domestic corporate securities and government bonds, as well as to fund second-period consumption.

Following the approach used by Peter Diamond and James Mirrlees (1971), the government, in setting its policy, is implicitly choosing values for w^n, t, g, K, π, and the method of corporate finance, allowing G to adjust to satisfy the resource constraint. These values are under the control of the government, given the available tax instruments, and together they determine the equilibrium for the economy. In setting its policy, the government must take as given the behavior of residents, which determine C_1, C_2, and L as functions of w^n and q.

In a closed economy, the characteristics of an optimal tax policy are normally rather complicated.[29] For this small open economy, however, the situation is quite different. First, whatever the optimal choice for K, this investment should be financed by equity. The investment rate is determined by q^c, but the cost of financing this investment depends on p^c. For any given value of q^c, p^c should be minimized and, as is seen in Figure 1, the minimum cost source of finance is equity. To insure that corporations do not use debt finance, the government could, for example, disallow interest deductions and deductions of capital losses on principal repayments, so that payments to both debt and equity holders would be treated the same for tax purposes.

More important, each country will find it optimal to set its corporate tax rate to zero. Intuitively, since the supply of capital from abroad to the economy is infinitely elastic, labor bears the entire burden of either a labor income tax or a corporate income tax, so both lead to a change in labor supply decisions. A corporate tax, however, simultaneously creates an additional distortion which reduces capital investment in the economy. It therefore is dominated by a labor-income tax. To demonstrate the result formally, differentiate the utility function, $V(w^n, q; Y, G)$, with respect to K allowing G to be determined by equation (2), while holding the other policy instruments, w^n, t, g, and π, constant. The resulting first-order condition states simply that $f_K = p^c$, which is satisfied whenever $\tau = 0$. A small country should therefore not attempt to tax capital, regardless of the tax policies in other countries.[30]

---

[28] If domestic residents buy domestic securities, then $p(t)$ will simply equal the return earned.

[29] See, for example, Feldstein (1978), King (1980), or David Bradford (1980).

Another result which follows immediately is that, whenever some taxable bonds earn a higher before-tax rate of return than equity,[31] a small country would never choose such a high value for the tax on interest income, t, that its residents end up investing in corporate equity. To see this, assume to the contrary that its residents find equity to be the best available investment, given the tax rate t. Then marginal reductions in this tax rate have no effect on the equilibrium. But as the tax rate is lowered, at some point investors shift their portfolios entirely into bonds earning a higher before-tax rate of return. At this point, individuals receive the same after-tax return with the bonds as with equity, but the country receives a higher real before-tax return with bonds. Government revenues jump when the investors shift into taxable bonds, at no cost to investors, and so the government would find it advantageous to lower t at least to this point.

Someone has to own the equity, however. If all countries are small, and lower their interest income tax sufficiently to induce their residents to invest in taxable bonds, then the only equilibrium is one in which all bonds plus some equity are held by investors facing equal tax rates on interest income and capital gains, so that the real before-tax returns on all assets are the same. In this situation, no country has an incentive to lower its tax rate further.[32] Tax competition forces all countries to choose tax rates such that $t_j = g_j$, for any given value of g_j.[33]

However, if a country chooses optimal values for g and t simultaneously, then it would set $t = g$ whenever the real before-tax returns differ on different assets, regardless of the tax policy in other countries.[34] To see this, consider what happens when a country raises g while reducing t as needed to keep q unchanged. These changes cause the tax parameter α to increase, inducing residents to shift their portfolio toward assets earning a higher before-tax return. These changes therefore increase government revenue at no cost to residents. By raising g, the government reduces the tax incentive investors face to shift their portfolios into assets earning a lower before-tax return.

The U.S. Treasury recently proposed a change similar to this when it recommended that exchange rate capital gains and losses on foreign bonds be made fully taxable. Under this proposal, with $g_j = t_j$, after-tax income on an investment in bonds from country k would equal $(1-t_j)v_k(r_k - \pi_k) - t_j v_j \pi_j$. An investor in bonds would now purchase the bond paying the highest before-tax real return, $v_k(r_k - \pi_k)$. However, since the tax treatment of equity would remain different, investors may still prefer to invest in equity even if its before-tax rate of return is lower.

---

[30] I ignore here any complications that arise if the corporation being taxed is a foreign-based multinational which owes further taxes when profits are repatriated. Bruce (1986) shows in this setting that if the firm's home country gives a tax credit for corporate taxes already paid, as does the United States, then the local country should tax the multinational at a rate equal to that in the home country.

[31] If enough countries tax interest income and capital gains at equal rates, so that all bonds plus some equity end up being owned by residents of these countries, then all securities will earn the same before-tax rate of return.

[32] If any set of countries choose values for t below g, then interest rates would adjust so that their investors earn a lower before-tax return on their savings. Given that the subsidy also uses revenue, these countries would face an incentive to raise t back up toward g.

[33] Similar interactions among the tax policies of different governments have been described in the fiscal federalism literature. See, for example, John Wilson (1985) or myself (1983).

[34] I ignore here any of the administrative problems of taxing gains at accrual.

The factors affecting the optimal common value of t and g are just the same as those which would appear when analyzing a closed economy. In general, as shown in the Appendix, when indirect utility $V(w^n, q; Y, G)$ is maximized with respect to the common value of t and g, allowing G to adjust according to the economy's budget constraint, the first-order condition for the optimal common value of t and g is

$$(3) \quad dV/dt = (Y - C_1)(\lambda - \theta)p + \lambda p[(p-q)S(C_1, q) - swS(L, q)] = 0.$$

Here, $\lambda = (\partial V/\partial G)/(1+i)$, which is the marginal utility received from the expenditures resulting from extra government revenues. In contrast, θ represents the marginal welfare that results when the individual receives an extra unit of resources in the second period, taking into account not only the utility the individual receives from the extra consumption, but also the value of the extra tax revenue received as a result of these extra expenditures.[35] Anthony Atkinson and Joseph Stiglitz (1980, p. 373) show that, at the optimum, $\lambda > \theta$ as long as tax revenues are positive. Finally, $S(L, q)$ and $S(C_1, q)$ represent the compensated effects of a change in q on L and C_1, respectively.

In this equation, the first term describes the change in the government's revenue from the tax on the return to savings, holding the amount of savings constant. When the government collects more revenue as the tax rate rises, there is some net benefit since alternative sources of revenue are also costly. The last two terms describe the compensated change in savings and labor supply resulting from the drop in the individual's net return to savings, the types of efficiency costs from tax increases focused on in past studies.[36] Under normal assumptions, the optimal common value of t and g would be positive. To show this, all we need to do is to demonstrate that $dV/dt > 0$ when evaluated at $t = 0$. The first term is clearly positive at $t = 0$, while the second term is clearly zero. Given the presumption that labor supply is unresponsive to interest rate changes, so that the last term is small, the entire derivative must be positive.

When in equilibrium different assets earn different before-tax rates of return, extra considerations also enter into the optimal choice for the domestic inflation rate. Given that the values of t and g are chosen optimally, the model implies that a country always has an incentive to lower its inflation rate. By lowering its inflation rate, a country lowers the *real* rate of return that it must pay on government debt, a clear gain. When a country lowers its inflation rate, it also lowers by a uniform amount the taxes collected on the returns from every asset that the resident might buy. However, this effect can be precisely offset by a suitable increase in the common values of t and g. See the Appendix for a formal proof of this argument.

B. *Other Policy Responses in an Open Economy*

If, for whatover reason, enough countries tax interest income more heavily than capital gains, so that there is a diversity of real before-tax returns on different financial assets in equilibrium, then governments face a variety of other peculiar incentives.

To begin with, while individuals and corporations have been assumed to be unable to engage in arbitrage, governments may not face the same constraints. Given the presence of these different rates of return, each government would wish to borrow at a low real rate and reinvest at a high real rate. For example, a low inflation rate country could issue debt to finance the purchase of foreign bonds issued by a high inflation rate country. Similarly, a high inflation country could borrow in a foreign currency (for example, debt denominated in dollars), and use the funds to retire any debt issued in its own currency.[37] Within the model, there is no natural limit to these arbitrage opportunities. Presumably such arbitrage is limited by the moral hazard problem of explicit default, or implicit default through unexpected inflation.

---

[35] Mathematically, $\theta = \partial U/\partial C_2 + \lambda[sw(\partial L/\partial Y) - (p-q)(\partial C_1/\partial Y)]/(1+q)$.

[36] See, for example, Bradford.

[37] This provides an intriguing explanation for the heavy indebtedness to U.S. banks of a number of developing countries with high domestic inflation rates.

Similarly, countries borrow at one real rate, while their investors save facing another real rate. When these two rates differ, there is an additional consideration when comparing debt vs. tax finance of government expenditures.[38] If a country has both a low inflation rate and a high value of α, then it faces an incentive to use debt finance, since it must pay a low real rate on its debt, whereas its investors, who increase their savings to cover future tax obligations, earn a higher before-tax real rate on their extra savings.

In contrast, a country with both a high inflation rate and a low value of α has an incentive to avoid any debt and even to try to build up a stock of foreign reserves. Its investors earn a lower real return on their savings than the government must pay on its debt. Such a government might also try to induce its investors to buy domestic bonds rather than foreign bonds. Simply outlawing the purchase of foreign assets, however, may lead investors to buy domestic equity rather than domestic debt, and so earn an even lower before-tax rate of return.

One other response a government might attempt to the restrictions capital mobility and tax competition impose on its tax policy is to negotiate with other governments concerning their joint tax policy. In isolation, each country sets its tax rates such that any small change has no effect on the utility of its representative citizen. However, a small change in one country's tax policy could well have first-order effects on the utility of nonresidents. When it does, there exists the possibility of mutually beneficial tax treaties.[39] Since an examination of the possibility of tax treaties inevitably involves examining decisions by a group of countries which collectively can affect market prices, a problem avoided elsewhere in this paper, I leave this topic to future research.

C. *Optimal Tax Policy with No International Trade in Equity*

One characteristic of the small open economy analyzed above is that tax rates on individual investors can be varied freely without affecting the required marginal product of corporate investment. This strong separation between the behavior of investors and the capital intensity of corporate production does not exist if there is no international trade in equity and no multinational firms, as in the model described at the end of Section I. Without this separation, additional complications arise when analyzing the effects of tax changes.

In the model with no international trade in equity, individual portfolios consist of both domestic equity and foreign bonds. The required return on equity, and so the amount of investment, now depend directly on the tax treatment of bonds. Whenever the taxes collected on a unit investment in capital differ from those collected on a unit investment in bonds, the portfolio composition of the country is distorted.

When the corporate tax rate is chosen optimally, this would not occur. Given the existing personal tax rates, the country faces a horizontal supply of funds at the price p, and ought to invest until $f_K = p$. This follows immediately if the individual's indirect utility function is differentiated with respect to K, taking into account how G is affected through the resource constraint of the economy. The latter is now $C_2 + (1+i)G = f(K, L) + K + (1+p)(Y - C_1 - K)$.

However, given the tax system, corporations choose to invest until $f_K(1-\tau) = q^c$. Therefore, at the optimal policy, $\tau = 1 - q^c/p$. If a domestic resident invests in bonds from some country m, then $p = v_m(r_m - \pi_m)$; if the corporation chooses to borrow in some country k, then $q^c = (1-d)v_j(e_j - \pi_j) + d[v_k(r_k(1-\tau) - \pi_k(1-\gamma)) - \gamma v_j \pi_j]$. In equilibrium, $v_j(e_j - \pi_j) = v_m(r_m(1-\alpha) - \pi_m)$. Substituting these equations for p and q^c, and using the result that $f_K = p$, we find that the optimal value of τ also satisfies

---

[38] Given the nature of the model used above, by tax finance I mean a lump sum tax imposed in the first period.

[39] See my article (1983) for a further discussion in the context of fiscal federalism.

$$(4)\quad \tau f K = (1-d)v_m r_m(1-\alpha) + d\left[f_K - \left(v_k\left(r_k(1-\tau) - \pi_k(1-\gamma)\right) - \gamma v_j \pi_j\right)\right].$$

Optimal tax policy still involves investing until the marginal product of capital equals the rate of return available on the world market, but if interest income is subject to tax, this is no longer accomplished by setting the corporate tax rate to zero. Instead, the tax system must be designed so that the same amount is collected in tax on an investment in corporate capital as is collected on an investment in foreign bonds. In equation (4), the left-hand side describes the corporate taxes that would be paid if the investment were equity financed. The first term on the right-hand side measures the extra taxes due on interest income from bonds since it, unlike capital gains income, is taxable at ordinary tax rates. The last term measures the corporate tax savings from use of debt finance.[40] The equation claims that the corporate tax rate should be set so that enough is collected in tax, given the tax savings from debt finance, to compensate for the heavier tax burden on interest income from foreign bonds.

The optimal corporate tax rate can easily be substantial. For example, if the corporation borrows domestically, if investors buy domestic bonds, and if $t=.30$, $r=.10$, $\pi=.04$, $g=.4t$, and $d=.25$, the optimal value of τ would be 0.44. Since the corporate tax is used to offset the more favorable personal tax treatment of capital gains income, this result is very sensitive to the tax treatment of capital gains. If instead $g=.2t$, then the optimal τ equals .58, while if $g=t$, the optimal τ equals zero.

The optimal value of the corporate tax rate can easily be much larger than the personal tax rate on interest income t, because the extra taxes due on nominal interest payments must be compensated for by an extra tax on *real* corporate income. In order to collect the same taxes in each case, the tax *rate* applied to the real income may need to be higher.

These results on optimal corporate taxation are in sharp contrast to those found in the previous section, where equity was assumed to be tradeable internationally. There, the optimal corporate tax rate was zero, regardless of the value chosen for t, whereas here the optimal corporate tax rate can be much higher than the value chosen for t. Optimal corporate tax policy thus depends critically on which assets are tradeable internationally.

III. Summary and Qualifications

This paper has explored the characteristics of individual portfolio holdings in a world economy with a unified securities market where there are many countries, each with its own tax rates and inflation rate. When nominal interest is fully taxable but income to equity owners is taxable only at lower capital gains rates in all countries, I show that the highest tax bracket investors specialize in equity and, among the remaining investors, those with lower tax rates buy bonds of countries with higher inflation rates.

Because of the tax system, governments of countries with a higher inflation rate must pay a higher *real* interest rate on their debt. This is necessary in equilibrium to compensate those who purchase the debt for their higher taxable income, and gives countries an incentive to lower their inflation rate.

Similarly, the combined rate of return received by investors and the government together on the savings undertaken by investors, which equals the real before-tax return, is higher in countries where the tax rates on interest income and capital gains are more similar. This occurs because investors facing relatively equal tax rates on all sources of income are less likely to buy a security earning a lower before-tax return because of its favorable tax treatment. Countries whose investors specialize in equity, because of a low tax rate on capital gains relative to other forms of income, receive the lowest real return on their savings. Optimal tax policy in a small country would involve equal tax rates on all forms of capital income, in order to avoid distorting the composition of individual portfolios.

---

[40] It is still the case, though, that optimal tax policy would eliminate the tax incentive favoring use of debt finance.

In this setting, a small country would not find it attractive to impose a corporate income tax. A tax on either capital or labor would be borne entirely by labor, but a tax on capital creates a further distortion to capital-labor ratios.

I also explore a model in which there is a unified world market in bonds, but no international trade in equity. Here, I find a strong tax incentive for firms owned by investors in countries with high personal tax rates to become multinationals and invest abroad. These firms require a low real return on their investments since their equity owners have a low opportunity cost of funds. It remains true, however, that the highest tax rate countries, which generate multinational firms, receive the lowest real return on their savings.

If multinationals are assumed not to play an important role, then investors in a country would own both bonds and domestic equity. In this case, the tax system distorts investors' portfolio choices whenever the total taxes paid on an equivalent investment in equity or foreign bonds differs. This is likely since nominal income on bonds is taxable under the personal tax, but only the real return on corporate capital is taxable under the corporate income tax. If equity investments are taxed more lightly, then there is an incentive on a government to raise its corporate tax rate, and/or lower its personal tax rate on interest income.

As with any study, these conclusions depend on the assumptions used, and a number of these assumptions are admittedly unrealistic. The most important one, perhaps, is the assumption of certainty. If asset returns were uncertain, then portfolios would become diversified, though would be pushed by the tax law in the direction described in this paper. In addition, as shown in my paper with Hal Varian, in the presence of uncertainty, even small countries will not normally be price takers, implying extra complications when analyzing optimal government policies. In particular, there will normally be incentives to restrict the amount of international trade in securities.

This paper has also ignored a number of other factors which limit the amount of international trade in securities. For example, I have ignored the existence of capital controls in some countries, and have ignored the threat of default or expropriation. In addition, I have ignored real exchange rate risk, which should push investors towards specializing in domestic securities.

I have also assumed that physical capital is fully mobile between countries, so implicitly examine a long-run equilibrium. In addition, I ignore the effects of a delayed response in capital reallocation on optimal tax policy.

Adding extra complications to the model undoubtedly would make the conclusions more complicated. The objective of this paper is to point out factors which arise even in simpler models, which should remain important as further complications are added.

APPENDIX

The objective of this Appendix is to derive the effects of a marginal change either in the joint values of g and t or in π on the welfare of the representative resident. The resident's welfare can be expressed, using his (her) indirect utility function, by $V(w^n, q; Y, G)$. Government expenditures must satisfy the resources constraint for the economy described in equation (2). Differentiating utility with respect to t, keeping $g = t$, holding the other policy instruments w^n, K, and π, constant, and allowing G to adjust according to the resources constraint implies

$$\text{(A1)}\quad dV/dt = V_2(\partial q/\partial t) + \frac{V_4}{1+i}\left[f_L\frac{\partial L}{\partial t} - (1+p)\frac{\partial C_1}{\partial t} - \frac{\partial C_2}{\partial t}\right].$$

Substituting for C_2, using the individual's budget constraint $C_2 = (Y - C_1)(1+q) +$

w^nL, and rearranging gives

$$\text{(A2)}\quad \frac{dV}{dt}=(V_2-\lambda(Y-C_1))\frac{\partial q}{\partial t}+\lambda\left(sw\frac{\partial L}{\partial q}-(p-q)\frac{\partial C_1}{\partial q}\right)\frac{\partial q}{\partial t}.$$

However, the analogy to Roy's identity in this context implies that $V_2=(Y-C_1)\times(\partial U/\partial C_2)$. In addition, we can use the Slutsky equation to decompose the effects of q on both L and C_1 into substitution effects and income effects. For example, $\partial L/\partial q=S(L,q)+[(Y-C_1)/(1+q)](\partial L/\partial Y)$. Finally, when $t=g$, $\partial q/\partial t=-p$. Substituting these equations and combining terms gives equation (3) in the text.

Following the same steps, it is easy to show that when the utility of the representative resident is differentiated with respect to π, given the budget constraint and the values of the other policy instruments, the resulting expression satisfies

$$\text{(A3)}\quad dV/d\pi=[V_2-\lambda(Y-C_1)](\partial q/\partial\pi)+\lambda\left[sw\frac{\partial L}{\partial q}-(p-q)\frac{\partial C_1}{\partial q}\right]\frac{\partial q}{\partial\pi}-\lambda G\frac{\partial i}{\partial\pi}.$$

Using equation (A2) to eliminate the first three terms in this equation implies that $\partial V/\partial\pi=-\lambda G(\partial i/\partial\pi)$.

REFERENCES

Adler, Michael and Dumas, Bernard, "International Portfolio Choice and Corporation Finance: A Synthesis," *Journal of Finance*, June 1983, *38*, 925–84.

______ **and** ______, "The Long-Term Financial Decisions of Multi-National Corporations," in Edwin J. Elton and Martin J. Gruber, eds., *International Capital Markets*, New York: American Elsevier, 1975, 360–87.

Atkinson, Anthony B. and Stiglitz, Joseph E., *Lectures on Public Economics*, New York: McGraw-Hill, 1980.

Auerbach, Alan J., "Wealth Maximization and the Cost of Capital," *Quarterly Journal of Economics*, August 1979, *93*, 433–46.

Boadway, Robin, Bruce, Neil, and Mintz, Jack, "Taxation, Inflation, and the Effective Marginal Tax Rate on Capital in Canada," *Canadian Journal of Economics*, February 1984, *17*, 62–79.

Bradford, David F., "The Economics of Tax Policy toward Savings," in G. M. von Furstenberg, ed., *The Government and Capital Formation*, Lexington: Ballinger, 1980, 11–71.

Bruce, Neil "Capital Income Tax Reform in a Capital Importing Economy," mimeo, February 1986.

Diamond, Peter and Mirrlees, James, "Optimal Taxation and Public Production, I: Production Efficiency (II: Tax Rules)," *American Economic Review*, March 1971 (June 1971), *61*, 8–27 (261–78).

Dooley, Michael P. and Isard, Peter, "Capital Controls, Political Risk and Deviations from Interest Rate Parity," *Journal of Political Economy*, April 1980, *88*, 370–84.

Feldstein, Martin S., "The Welfare Cost of Capital Income Taxation," *Journal of Political Economy*, April 1978, *86*, S29–S52.

______ **and Hartman, David,** "The Optimal Taxation of Foreign Source Investment Income," *Quarterly Journal of Economics*, November 1979, *93*, 613–29.

Fraga, Arminio, "The Exchange Rate Risk Premium and the Micro-Foundations of Asset Markets Models of the Exchange Rate," Woodrow Wilson School Discussion Papers in Economics No. 83, November 1984.

Frisch, Daniel, "Issues in the Taxation of Foreign Source Income," in Martin S. Feldstein, ed., *Behavioral Simulation Methods in Tax Policy Analysis*, Chicago: University of Chicago Press, 1983.

Galper, Harvey and Toder, Eric, "Transfer Elements in the Taxation of Income from Capital," in Marilyn Moon, ed., *Economic Transfers in the United States*, Chicago: University of Chicago Press, 1984.

Gordon, Roger H., "Inflation, Taxation, and Corporate Behavior," *Quarterly Journal of Economics*, May 1984, *99*, 313–27.

______, "Interest Rates, Inflation, and Corporate Financial Policy," *Brookings Papers on Economic Activity*, 2:1982, 461–88.

______, "An Optimal Taxation Approach to Fiscal Federalism," *Quarterly Journal of Economics*, November 1983, *98*, 567–86.

______ **and Varian, Hal** "Taxation of Asset Income Given the Presence of a World Securities Market," mimeo., July 1986.

Goulder, Lawrence H., Shoven, John B. and Whalley, John, "Domestic Tax Policy and the Foreign Sector," in Martin S. Feldstein, ed., *Behavioral Simulation Methods in Tax Policy Analysis*, Chicago: University of Chicago Press, 1983, 333–64.

Hartman, David G., "On the Optimal Taxation of Capital Income in the Open Economy," NBER Working Paper No. 1550, January 1985.

______, "Taxation and the Effects of Inflation on the Real Capital Stock in an Open Economy," *International Economic Review*, June 1979, *20*, 417–26.

______, "Tax Policy and Foreign Direct Investment," NBER Working Paper No. 689, June 1981.

King, Mervyn A., *Public Policy and the Corporation*, New York: Wiley & Sons, 1977.

______, "Savings and Taxation," in G. A. Hughes and G. M. Heal, eds., *Public Policy and the Tax System*, London: Allen & Unwin, 1980.

Kyle, Albert S., "Market Structure, Information, Futures Markets, and Price Formation," in Gary G. Storey et al., eds., *International Agricultural Trade: Advanced Readings in Price Formation*, Boulder: Westview, 1984, 45–64.

Levi, Maurice, "Taxation and 'Abnormal' International Capital Flows," *Journal of Political Economy*, June 1977, *85*, 635–46.

Miller, Merton H., "Debt and Taxes," *Journal of Finance*, May 1977, *32*, 261–75.

Makin, John H., "Exchange Rates and Taxes," NBER Working Paper No. 1350, May 1984.

Tanzi, Vito "Inflationary Expectations, Economic Activity, Taxes, and Interest Rates," *American Economic Review*, March 1980, *70*, 12–21.

______, *Taxation, Inflation, and Interest Rates*, Washington: International Monetary Fund, 1984.

Wilson, John D., "Trade, Capital Mobility, and Tax Competition," Social Systems Research Institute, Workshop Series No. 8505, University of Wisconsin, 1985.

The Dynamics of Population Growth, Differential Fertility, and Inequality

By DAVID LAM*

This paper analyzes the effects of income differentials in fertility on Lorenz curves and standard inequality measures. The role of intergenerational mobility is examined and incorporated into counterfactual simulations based on Brazilian data. Two standard inequality measures move in opposite directions in both the steady state and the transition in response to the elimination of fertility differentials. The counterfactuals confirm the theoretical predictions of misleading intertemporal inequality comparisons in the presence of differential fertility.

The relationship between population growth and the distribution of income has been investigated by development economists and demographers for the last two decades. Population growth rates have been included as explanatory variables in cross-national empirical investigations of inequality, with researchers concluding from both single-equation studies, such as Irma Adelman and Cynthia Taft Morris (1973), M. S. Ahluwalia (1976), and Rati Ram (1984), and from simultaneous equation models, such as Robert Repetto (1979) and C. R. Winegarden (1978, 1980), that higher population growth rates increase inequality.[1] Unfortunately, these empirical results are disturbingly sensitive to model specification and the definition of sample countries. Although the case for simultaneous equation models is convincing, the identifying restrictions are not. Even if a consistent empirical relationship between population growth and inequality were established, the policy implications would be unclear. The commonly cited reasons for effects of population growth on inequality include significant welfare issues, such as declines in wages relative to rents on capital and land, and issues of minimal welfare significance, such as pure compositional effects of a younger age structure.

One issue which raises especially difficult methodological and welfare questions is the effect on inequality of differential fertility across income classes. Ahluwalia suggests that "the most important link between population growth and income inequality is provided by the fact that different income groups grow at different rates, with the lower income groups typically experiencing a faster natural rate of increase" (p. 326). The existence of differential fertility, however, makes the relationship between population growth and inequality more complex than existing empirical literature has recognized. Given differential fertility across income groups, compositional effects will confound observed relationships between population growth and inequality. As demonstrated below, compositional effects on measures of dispersion are more difficult to identify than compositional effects on mean values, and can easily lead to misleading inferences from cross-national or time-series comparisons of inequality measures.

Section I demonstrates the sensitivity of standard inequality measures to changing population composition, independent of changes in relative incomes. The tendency

---

*Department of Economics and Population Studies Center, University of Michigan, Ann Arbor, MI 48109. This research has benefited from discussions with Theodore Bergstrom, Chin-Yi Chu, and Lowell Taylor, and the research assistance of Deborah Levison. The cooperation of Tom Merrick in providing the 1976 Brazilian PNAD survey and generously sharing his knowledge of the data set is gratefully acknowledged.

[1] A recent exception is Gerry Rodgers (1983), who fails to find a statistically significant direct effect of population growth on inequality. For surveys of the literature on the distributional effects of population growth, see Bryan Boulier (1977), Rodgers, and my article (1986a).

for inequality measures to provide conflicting signals in the presence of differential fertility is further established in Section II, which demonstrates an inherent tendency for differential fertility to produce Lorenz curves which intersect from period to period. Section III discusses the role of mortality, age structure, and intergenerational mobility, and incorporates them into the analysis through the use of a multistate life table. Section IV documents the pattern of differential fertility in Brazil, and uses the data to construct a multistate life table to project the Brazilian distribution of income. The results of two counterfactual simulations indicate that an increase in the fertility of the poor in Brazil would increase the coefficient of variation, but would decrease the variance of log income, in both the steady state and the transition. The contradictory results for the two measures are consistent with the theoretical analysis, and confirm the difficulty of inferring changes in relative welfare from changes in inequality indexes in the presence of differential fertility.

I. The Mathematics of Population Growth and Inequality

Recent literature on the measurement of inequality has analyzed the consistency between inequality measures and social welfare functions, clarifying the value judgements implicit in particular inequality measures.[2] Much less attention has been given to the effects on measures of inequality of introducing new members to the population.[3] If the population is expanded differentially by income level, the resulting effects on inequality measures need not be consistent with the social welfare function implicit in a particular index. The problems can be demonstrated by analyzing the effect of some arbitrary expansion of the population on the variance of the logarithm of income, a standard inequality measure.[4] Consider an initial population of n persons with mean of log income μ_n and variance of log income σ_n^2 to which we add a new group of m persons with log mean μ_m and log variance σ_m^2.[5] The log mean will rise or fall as the entering group's mean is above or below the mean of the original population, a simple compositional effect which depends only on the difference in the two groups' mean incomes and is affected in magnitude but not in direction by the size of the entering group.

Expressing the combined log variance, σ_{n+m}^2, as a standard decomposition of between-group and within-group variances, the new log variance reduces to

$$(1)\quad \sigma_{n+m}^2 = \frac{nm}{(n+m)^2}(\mu_m - \mu_n)^2 + \frac{n}{n+m}\sigma_n^2 + \frac{m}{n+m}\sigma_m^2.$$

The new log variance is a simple function of the two population sizes, the two original log variances, and the difference between the two original log means. If differential fertility by income levels produces an entering population that is poorer than the original population, will inequality, as measured by log variance, increase or decrease? It is clear from (1) that any result is possible, the direction of the change depending on all of the parameters in (1). One set of implications can be summarized by noting the following condition for the change in the log variance:

---

[2] See the seminal article by Anthony Atkinson (1970). Subsequent contributions are surveyed by Nanak Kakwani (1980).

[3] Exceptions include Sherman Robinson (1976) and Gary Fields (1979), who analyze the effects of changing the sectoral composition of the population, and Samuel Morley (1981), who points out a number of problems underlying inequality comparisons in the presence of population growth.

[4] The mathematics are formally similar to the effects of transferring individuals across income classes, since any changes in the probability density function will cause changes in measures of dispersion. Subramian Swamy (1967), Robinson, Fields, and Oded Stark and Shlomo Yitzhaki (1982) have shown that transferring individuals from a low-wage rural sector to a high-wage urban sector can generate U-shaped paths in measures of inequality over time. As emphasized by Fields and Stark-Yitzhaki, the welfare implications of this urbanization process may be much different than inequality measures imply, since the transfer of individuals into higher-income classes is an unambiguous Pareto improvement, in spite of the increase in inequality which standard inequality measures will indicate during some portion of the transition. Welfare interpretations of changes in fertility patterns are even more ambiguous, since a welfare ranking of alternative fertility scenarios requires comparing distributions with different numbers of persons, some of whom do not exist *ex ante*.

[5] Log variance and log mean will be used to refer to the variance and mean, respectively, of the natural logarithm of income.

$$(2)\quad \sigma^2_{n+m}\{\lesseqgtr\}\sigma^2_n \quad \text{as } m\{\gtreqless\}\frac{n(\mu_m-\mu_n)^2}{\sigma^2_n-\sigma^2_m}-n.$$

If the entering group's log variance is smaller than the original log variance and the new group's log mean is more than one standard deviation from the original log mean,[6] then there exists some number of new entrants which will leave the original log variance unchanged. The value of m which will maintain the original log variance increases as the new entrants' mean moves farther from the original mean. We cannot only satisfy the equality in (2) by an appropriate choice of m, we can also satisfy either of the inequalities, either increasing or decreasing log variance as we change the size of the entering group. We thus see the more complex nature of compositional effects on measures of dispersion compared to compositional effects on means. While a group of low-income entrants will unambiguously decrease mean income when added to a higher-income population, the direction of the effect on the log variance cannot be predicted without knowing all of the parameters in (2).

The log variance is analytically convenient because of its simple decomposition properties. Similar results can be derived for all additively decomposable measures. The analog to (2) for the squared coefficient of variation, for example, is

$$(3)\quad C^2_{n+m}\{\lesseqgtr\}C^2_n \text{ as } m\{\gtreqless\} \frac{n\left(\bar{y}^2_m C^2_m+\bar{y}^2_n C^2_n-2\bar{y}_m\bar{y}_n C^2_n\right)+n\left(\bar{y}_m-\bar{y}_n\right)^2}{\bar{y}^2_m\left(C^2_n-C^2_m\right)}$$

where C_i is the coefficient of variation and $\bar{y}_i$ is the arithmetic mean for group i. A comparison of (2) and (3) reveals that there is no necessary relationship between the number of entrants m necessary to hold the log variance constant and the number of entrants necessary to hold the coefficient of variation constant. A given size group of entrants could increase inequality by one measure but decrease it by the other measure. Indeed, since it is possible to have $\sigma^2_n > \sigma^2_m$ at the same time that $C_n < C_m$, it may be possible to satisfy the equality in (2) but impossible to satisfy the equality in (3) for any positive m.

These results have important implications for the patterns of inequality that will be observed whenever income classes grow at different rates. If one income class begins to grow at a faster rate than the rest of the population, the log variance of income will change by the dynamics implied by (1). If the class has a mean of log income more than one standard deviation from the original log mean and a within-class variance of log income smaller than the original log variance, a small increase in its relative size will tend to increase log variance. If the group continues to grow, the log variance will eventually decline, returning to its original level and converging in the limit to the intragroup log variance of the rapidly growing class. Analogously, the coefficient of variation will change according to the conditions in (3). The change in the coefficient of variation need not be in the same direction in any period as the change in the log variance, but the coefficient of variation must also eventually decline as long as $C_m < C_n$.[7] The fact that these two standard inequality measures may move in opposite directions in response to differential fertility is the result of a general tendency for ambiguous inequality comparisons which can be demonstrated by analyzing the effect of differential fertility on Lorenz curves.

---

[6] The restrictions are assumed in order to limit the analysis to positive values of m, i.e., to additions rather than subtractions from the population. The mathematical result in (2) is more general, but will imply $m < 0$ if $|\mu_m - \mu_n| < \sigma_n$ or if $\sigma^2_m > \sigma^2_n$.

[7] For measures which are not additively decomposable, such as the Gini coefficient, the effect of new entrants with some within-group inequality cannot be

II. The Effect of Population Expansion on Lorenz Curves

The Lorenz curve, which expresses cumulative shares of income as a function of cumulative shares of the population,[8] has long served as a foundation for inequality comparisons. The effects of differential fertility on inequality are clarified by identifying the conditions under which individuals can be added to a population such that the Lorenz curve for the combined population lies above the Lorenz curve for the original population. Since inequality comparisons are inherently ambiguous when Lorenz curves intersect, establishing the necessary and sufficient conditions for Lorenz dominance provides additional insights into the problems with inequality comparisons under differential fertility.

Consider an initial population with an income distribution described by the density function $f_1(y)$ to which we add a second population with income distribution $f_2(y)$. Under what conditions will it be the case that the combined distribution, $f_3(y)$, Lorenz dominates the original distribution? The conditions are surprisingly restrictive: If $y_1^{\min} > 0$, where $y_1^{\min}$ is the income of the poorest individual included in the income distribution, then

$$(4) \quad \Phi_3(F) \geq \Phi_1(F)\forall F \quad \text{iff} \quad \mu_2 = \mu_1$$

$$\text{and} \quad \Phi_2(F) \geq \Phi_1(F)\forall F,$$

where $\Phi_i(F)$ is the Lorenz curve and μ_i is the mean for distribution $f_i(y)$. The combined population's income distribution will Lorenz dominate the original population's income distribution if and only if the entering population has the same mean as the original population and the income distribution of the entering population Lorenz dominates the original population's income distribution.

The most surprising result in (4), and the result most important for inequality comparisons under differential fertility, is the necessary condition that the entering population has the same mean income as the original population. Since the equal means condition will rarely be satisfied in practice, the result implies that the second-period distribution will either be unambiguously less equal than the original distribution, or that the Lorenz curves for the two periods will intersect.

A complete proof of (4) is contained in my paper (1986b), which analyzes the general properties of Lorenz curves produced from combinations of distributions. It is worth sketching the proof, however, since it demonstrates the problems inherent in inequality comparisons in the presence of changing population composition. The equal means condition results from the fact that Lorenz curves (like standard inequality measures) are independent of scale, depending only on relative income shares. Any change which reduces the cumulative income share of any cumulative share of the population will cause the new Lorenz curve to lie below the original Lorenz curve in that region.

Since by construction $\Phi_i(0) = 0$ and $\Phi_i(1) = 1$, in order for the Lorenz curve of the combined distribution to lie nowhere below

---

reduced to simple analytical expressions like (2) and (3) which depend only on measures of within-group inequality and within-group means. Although decompositions have been derived for the Gini coefficient (for example, C. H. Fei, Gustav Ranis, and Shirley Kuo, 1978), the within-group means and within-group Ginis are not sufficient statistics. More detailed data on within-group distributions or correlations across groups are also required. The logic clearly extends to any measure of dispersion, however, since the disproportionate growth of any group which has lower dispersion than the population as a whole must eventually cause total dispersion to decline as that group begins to dominate the distribution.

[8] For an income distribution with probability density function $f(y)$, the horizontal axis for the Lorenz curve is the cumulative distribution function $F(\hat{y}) = \int_0^{\hat{y}} f(y)\,dy$. The vertical axis is the first moment distribution function $\Phi(\hat{y}) = (1/\mu)\int_0^{\hat{y}} yf(y)\,dy$, where μ is the mean of the distribution.

the Lorenz curve of the original distribution, it must be true that $\Phi_3' \geq \Phi_1'$ as $F \to 0$ and that $\Phi_3' \leq \Phi_1'$ as $F \to 1$. The slope of the curve out of the origin is $y^{\min}/\mu$, where $y^{\min}$ is the income of the poorest person in the population, assumed to be nonnegative, and the slope coming out of the upper right-hand corner is $y^{\max}/\mu$, where $y^{\max}$ is the income of the richest person in the population.[9] Lorenz dominance of the combined distribution therefore requires that

$$(5) \qquad y_3^{\min}/\mu_3 \geq y_1^{\min}/\mu_1$$

$$\text{and} \qquad y_3^{\max}/\mu_3 \leq y_1^{\max}/\mu_1.$$

The addition of f_2 implies the introduction of new individuals to a population while leaving the incomes of the original members unchanged, so it must be true that

$$(6) \quad y_3^{\min} \leq y_1^{\min} \quad \text{and} \quad y_3^{\max} \geq y_1^{\max}.$$

But by the inequality conditions in (6), the slope conditions in (5) cannot both be satisfied for any $\mu_3 \neq \mu_1$ as long as $y_1^{\min} > 0$.[10] If $\mu_3 > \mu_1$ (implying $\mu_2 > \mu_1$), (6) requires that $y_3^{\min}/\mu_3 < y_1^{\min}/\mu_1$, implying that Φ_3 must lie below Φ_1 out of the origin. By the same logic, if $\mu_3 < \mu_1$, the new Lorenz curve must lie below the original Lorenz curve as $F \to 1$.[11] It follows that if $\mu_3 \neq \mu_1$, it must be the case that $\Phi_3(F) < \Phi_1(F)$ over some range of F.

We see, then, that in order for the combined distribution to Lorenz dominate the initial distribution, it is a necessary condition that $\mu_1 = \mu_2$. Given equal means, the second necessary condition in (4), which states that the entering population's distribution must Lorenz dominate the original distribution, is proven by showing that Anthony Atkinson's theorem on the equivalence of Lorenz dominance and second-degree stochastic dominance for distributions with equal means implies that

$$(7) \qquad \Phi_3(F) \geq \Phi_1(F) \forall F$$

$$\text{iff} \quad \Phi_2(F) \geq \Phi_1(F) \forall F.$$

Having shown that it is a necessary condition that the two populations have the same mean, (4) is established.

The result in (4) is a powerful statement of an inherent tendency for changes in population composition to produce ambiguous inequality comparisons. It applies not only to cases in which all members of the original population retain their initial incomes, but to any change in the population that satisfies (6), implying that the minimum income in the population does not increase and the maximum income in the population does not decrease, regardless of what happens elsewhere in the distribution. The result is easily extended to the effects of new entrants into a population with rising incomes. In such a case, (4) is modified to require that any group entering the population in period t must have a mean income of $\mu_1(1+g)^t$, where μ_1 is the mean income of the original

---

[9] In general, the slope of the Lorenz curve in discrete terms is

$$\frac{\Phi[F(y_k)] - \Phi[F(y_{k-1})]}{F(y_k) - F(y_{k-1})} = \frac{y_k}{\mu},$$

where y_k is the income of the kth person in the rank-ordered distribution.

[10] If $y_1^{\min} = 0$ in the original population, then it is possible to raise the mean without lowering the share of income to the poorest individuals in the population, since their shares are already at zero. This introduces additional conditions which will produce Lorenz dominance, involving Lorenz curves which initially coincide with the horizontal axis. The result is similar to the perverse properties Lorenz curves exhibit when negative incomes are included. In this case, I extend the usual restriction against negative incomes to exclude incomes at exactly zero as well.

[11] Note that if the extreme incomes remain constant, a change in the mean must imply that Φ_3 intersects Φ_1. Since the slopes at the extreme points are simply the extreme incomes divided by the mean, the slopes at both $F \to 0$ and $F \to 1$ must change in the same direction if $y^{\min}$ and $y^{\max}$ remain constant. But since Φ_1 and Φ_3 are equal at $F = 0$ and $F = 1$, both curves cannot increase (decrease) their slopes at the endpoints unless they intersect. If either $y_3^{\min}$ or $y_3^{\max}$ lies outside the range $[y_1^{\min}, y_1^{\max}]$, it is possible for the new Lorenz curve to lie everywhere below the original Lorenz curve.

population in the initial period, and g is the rate of growth of the original members' incomes. The requirement that the new group's income distribution Lorenz dominate the original group's income distribution is unchanged.

III. Mortality, Age Structure, and Intergenerational Mobility

The results above do not give a complete picture of the effects of differential fertility, since they ignore exits from the population due to mortality, do not account for the ages of new entrants, and do not allow for mobility across income groups. This section demonstrates that these issues are readily incorporated into a model which projects both age structure and the distribution of income using a multistate life table. Before combining the effects into a single model, however, a number of points can be made about each issue.

Note first that one method of incorporating mortality (or retirements, if the relevant population is economically active adults) is to define the group of "new entrants" in Sections I and II as actual entrants minus departures. This interpretation is straightforward as long as the density function for net entrants is nonnegative at all incomes, that is, as long as there are at least as many entrants as departures at every income level. If departures result from advancing age, as in deaths or retirements, then entrants will normally exceed departures in a growing population, allowing the analysis to carry through directly.[12] This interpretation is more problematic if there are more departures than entrants at any income level, since this implies a density function for new entrants which is negative at some incomes. The analysis must be modified in such a case, since such a density function must be ruled out for both the calculations of inequality measures for new entrants in Section I and the derivation of the Lorenz curve for new entrants in Section II.

A related issue is the treatment of age structure. New entrants into the economically active population enter at a young age, with low incomes, an issue which is not considered explicitly in the results of the previous sections. Although the results above can clearly be applied to the case of new young workers entering the labor force, I do not want to confound the effects of differential fertility with the pure compositional effects of age structure.[13] It does not appear appropriate to use the approach of Samuel Morley (1981), who in comparing Brazilian inequality in 1960 and 1970 separates the "base period poor" from new entrants by removing first-time members of the labor force from the 1970 distribution. This implies tracking the base-period poor as they advance ten years along their age-earnings profiles. While this may be interesting in its own right, it is primarily an indicator of the properties of age profiles of earnings and inequality and should be expected to show significant changes even when all life cycle earnings profiles remain constant between periods.

Introducing intergenerational mobility into the analysis is important because no matter how great the fertility differentials across income classes, a steady-state distribution of income will exist whenever the children of all income classes have some positive probability of entering any other income class. Introducing even a small degree of mobility is an important modification, since it means that persistent higher fertility for the poor will not cause the population to converge to 100 percent poor.[14]

---

[12] This will necessarily be true in a stable population with a positive growth rate, but need not be true in a population which has experienced a change in vital rates and has not converged to a new stable population, even with a positive growth rate.

[13] For an analysis of the effects on inequality of fertility-induced changes in age structure, see my article (1984).

[14] See Samuel Preston's (1974) analysis of the interaction of occupational mobility and occupation differentials in fertility. See also Robinson and Kermel Dervis (1977), who discuss the relationship between mobility and the distribution of income, and Chin-Yi Chu (1984), who develops a model of income distribution with endogenous fertility and bequest behavior.

Consider the following model of differential fertility and intergenerational mobility based on a simple Markov process governing transitions across income classes. Assume a vector $\mathbf{P}_t$ denotes the size of n income classes in period t. Class-specific fertility rates are described by a diagonal matrix of net reproduction rates, **F**, and intergenerational mobility is described by a matrix **M**, where element M_{ij} is the probability that a child of class j becomes a member of class i. Thinking of each period as a generation, the distribution of income in period t is

$$(8) \qquad \mathbf{P}_t = \mathbf{MFP}_{t-1}.$$

The properties of (8) are determined by the product of the income class transition matrix and the fertility matrix, which produce a combined transition matrix $\mathbf{T}^* = \mathbf{MF}$. The process governed by $\mathbf{T}^*$ will converge to a stable proportional income distribution if $\mathbf{T}^*$ is an irreducible primitive matrix.[15] Ignoring mortality and age structure for the moment, the effects of differential fertility on both steady states and transitions can be seen in the dynamics of period-to-period changes in the relative size of the poorest class. Denote the size of income class i at time t by $P_{i,t}$ and the proportion of the population in income class i at time t by $\pi_{i,t}$. Note that the proportion of the population in the poorest class, $i = 1$, at time t is

$$(9) \quad \pi_{1,t} = \sum_{i=1}^{n} P_{i,t-1} F_i M_{1i} \Big/ \sum_{i=1}^{n} P_{i,t-1} F_i.$$

Considering the effect of a change in F_1, the net reproduction rate of the poorest class, on the proportion in that class in the next period, the result in elasticity form reduces to

$$(10) \quad \partial \ln \pi_{1,t} / \partial \ln F_1 = M_{11} - \sum_{i=2}^{n} \gamma_{i,t} M_{1i},$$

where $\gamma_{i,t} = P_{i,t-1}F_i / \Sigma P_{i,t}$ is the proportion of the th period population born to parents from the ith income class. The elasticity of the proportion of the population in the poorest class with respect to the reproduction rate of that class is equal to the difference between the probability that a child of poor parents remains poor and the weighted average of the probabilities that children of all other classes become poor, with weights equal to the proportion of the period t population born to parents of each class.

From (10) we see that if there exists any $M_{1i} > M_{11}$, implying that parents in at least one class are more likely to produce poor offspring than are the poor themselves, then there will exist some initial income distribution (implying a set of weights γ_i) for which an increase in the fertility of the poor will cause a *decrease* in the proportion poor in the second period. If $M_{1i} \leq M_{11} \forall i$, then an increase in the fertility of the poor will always lead to an increase in the proportion poor in subsequent periods. Since the income distribution and hence the weights γ_i change during the transition to the steady state, (10) also implies that the effect of higher fertility among the poor may be different in every period until the ergodic income distribution is achieved.[16] The result in (10) describes the effect of higher fertility for the poor on the steady-state distribution of income if we replace $\pi_{i,t}$ with π_i^*, the steady-state proportion in class i, and $\gamma_{i,t}$ with γ_i^*, the proportion of the steady-state population born to parents of income class i, both of which will be independent of t and the initial distribution of income once the ergodic distribution is attained. If $M_{1i} < M_{11} \forall i$, the case which is probably most plausible empirically, then an increase in the fertility of the poor will unambiguously increase the percent poor in the steady state. If $M_{1i} > M_{11}$ for some but not all i, then the effect of higher fertility for the poor on the steady-state percent poor cannot be predicted without solving for the steady-state values explicitly.

---

[15] The proportions of the population in each income class will converge to constant proportions, with the absolute numbers growing at the same constant exponential rate. Irreducibility requires that transitions can occur between all states in the system. The stronger condition of primitivity requires that some power of $\mathbf{T}^*$ has only positive elements. If all fertility rates and transition probabilities are positive, then this condition is satisfied, since $\mathbf{T}^*$ consists entirely of positive elements. A convenient discussion of these properties can be found in Nathan Keyfitz (1977, pp. 27–33); a proof is provided by Beresford Parlett (1970).

[16] Specifically, if and only if there exists at least one $M_{1i} > M_{11}$ and at least one $M_{1j} < M_{11}$, then there will exist some initial income distribution for which higher fertility among the poor will increase the percent poor in some periods but decrease the percent poor in other periods, when compared to the transition path implied by the original fertility rates. If M_{1i} is less than (greater than) M_{11} for all i, then the transition path implied by a higher fertility rate for the poor will have a higher (lower) percentage poor in every period than the original transition path.

The effect of a change in F_1 on some inequality index I is

$$\frac{\partial I}{\partial F_1} = \sum_{i=1}^{n} \frac{\partial I}{\partial \pi_i} \frac{\partial \pi_i}{\partial F_1}. \tag{11}$$

Even if the signs of the effects of F_1 on the proportion in each class are unambiguous, we know from Sections I and II that the effect on a given inequality index of changes in the density function cannot be predicted without knowing the actual magnitudes of the changes. Even if the signs of $\partial \pi_i / \partial F_1$ are the same for any given i in all periods during the transition, the signs of $\partial I / \partial \pi_i$, and hence the sign of $\partial I / \partial F_1$, need not be constant across periods. Furthermore, we know from Sections I and II that the direction of change in one index need not be the same as the direction of change in some other index, in either the steady state or any period of the transition.

The effects of mortality, age structure, and intergenerational mobility can be analyzed simultaneously with a multistate life table, a multidimensional generalization of a standard life table. Just as standard life tables describe transition probabilities across age groups, multistate life tables describe simultaneous transitions across age groups and other states.[17] Specifically, I want to model transitions across income classes and age groups, where in the most general case the combination of mortality and economic mobility produces a separate conditional probability $T_{i_k, j_{k+1}}$ for every transition from income class i at age k to income class j at age $k+1$. In the most general case, I also allow each income group at each age to have a separate fertility rate $F_{i,k}$.

The generalization of the process in (8), which has no age structure, to allow for n income classes in m age groups is given by

$$\begin{bmatrix} \mathbf{P}_{1,t} \\ \mathbf{P}_{2,t} \\ \vdots \\ \mathbf{P}_{m,t} \end{bmatrix} = \begin{bmatrix} \mathbf{M}_1 & 0 & 0 & 0 \\ 0 & \mathbf{M}_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mathbf{M}_m \end{bmatrix} \times \begin{bmatrix} \mathbf{F}_1 & \mathbf{F}_2 & \cdots & \mathbf{F}_m \\ \mathbf{S}_1 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & \mathbf{S}_{m-1} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{P}_{1,t-1} \\ \mathbf{P}_{2,t-1} \\ \vdots \\ \mathbf{P}_{m,t-1} \end{bmatrix} \tag{12}$$

where $\mathbf{P}_{i,t}$ is an $n \times 1$ vector giving the size of each income class in the ith age group in period t, $\mathbf{M}_i$ is an $n \times n$ mobility matrix describing transitions across income classes for persons entering the ith age group, $\mathbf{F}_i$ is an $n \times n$ diagonal matrix giving the income-specific fertility rates for persons aged i, and $\mathbf{S}_i$ is an $n \times n$ diagonal matrix giving the income-specific probabilities of surviving from age group i to age group $i+1$.

The generalization of the ergodicity result for the process without age structure in (8) is that the multistate process in (12) will converge to a steady-state population with constant proportions in each of the $n \times m$ age-income groups if the product of the mobility matrix and the matrix of fertility and mortality rates is an irreducible primitive matrix. If at least one age group has a mobility matrix that allows transitions across all income classes, plausible restrictions on the fertility submatrices will guarantee convergence. Specifically, a sufficient (but not necessary) condition for convergence is that fertility rates are positive for all income classes in two adjacent age groups.

---

[17] See the contributions in Kenneth Land and Andrei Rogers (1982) for both theoretical analysis and applications.

TABLE 1—MEASURES OF MEAN INCOME, INCOME INEQUALITY, AND FERTILITY BRAZILIAN MALE HOUSEHOLD HEADS, 1976

	Individual Income		Log Income		Sons Aged 0–15		Total Births	
Quintile (1)	Mean (2)	Coefficient of Variation (3)	Mean (4)	Variance (5)	Mean (6)	Standard Deviation (7)	Mean (8)	Standard Deviation (9)
Males Aged 25–30 ($N = 9451$)								
1	514	0.287	6.186	0.151	0.938	1.027	2.413	2.001
2	918	0.132	6.814	0.017	0.879	0.959	2.142	1.661
3	1469	0.140	7.283	0.020	0.816	0.898	1.913	1.585
4	2438	0.165	7.785	0.027	0.710	0.826	1.644	1.322
5	7821	1.485	8.753	0.314	0.588	0.771	1.312	1.174
Total	2632	2.221	7.364	0.866	0.786	0.909	1.886	1.621
Males Aged 40–45 ($N = 8610$)								
1	553	0.267	6.268	0.117	1.730	1.492	6.189	3.652
2	968	0.147	6.864	0.021	1.725	1.501	5.647	3.359
3	1640	0.158	7.390	0.026	1.546	1.428	5.065	3.078
4	2945	0.196	7.969	0.039	1.313	1.301	4.441	2.893
5	10991	1.242	9.049	0.383	1.064	1.114	3.449	2.292
Total	3420	2.115	7.508	1.028	1.475	1.399	4.947	3.228

IV. Empirical Evidence from Brazil

Given data on income- and age-specific fertility, mortality, and interclass mobility, (12) can be used to analyze the dynamics of income distribution and age structure from some initial conditions to the steady state. Although the data for a complete version of (12) would be difficult to assemble for any population, simplified versions are feasible to construct and can be quite instructive. Data from Brazil are used to construct a model that captures the demographic realism of (12), but is stripped down to the simplest elements necessary to test whether an increase in the fertility rates of the poor will increase or decrease income inequality. Income and fertility data are drawn from the 1976 PNAD[18] survey of Brazilian households to construct a model of the male population with five income classes and four fifteen-year age groups. Ergodicity is guaranteed by assuming that sons of fathers from all classes have a positive probability of entering every income class as adults, and by assuming that the second and third age groups have positive fertility rates for all income classes.

The income and fertility data used to construct the transition matrix are shown in Table 1.[19] Columns 6 and 8 show a simple pattern of differential fertility in Brazil, with the number of 0–15-year-old sons and the total number of births declining monotonically as income increases for both 25–30 and 40–45-year-old men. Men in the poorest quintile at age 40–45 have experienced an average of 6.2 total births, compared to only 3.4 for men in the richest quintile.[20] Table 1 also indicates that inequality within quintiles is highest in the upper quintile by both the log variance and the coefficient of variation, with inequality in the bottom quintile slightly higher than inequality in the middle of the distribution.[21]

---

[18] The Pesquisa Nacional de Amostra Domiciliar is a household survey conducted by the Instituto Brasileiro de Geografia e Estatistica. The 1976 PNAD is a stratified sample which includes 99,167 households. The results reported in Table 1 use the weights included in the data set by IBGE to produce a representative sample of Brazilian households. The sample sizes reported in Table 1 are the unweighted number of observations.

[19] The quintile-specific mean number of sons 0–15 for men 25–30 and men 40–45 are interpreted as 15-year fertility rates and become the diagonal entries of submatrices $\mathbf{F}_2$ and $\mathbf{F}_3$ in (12). It is assumed that $\mathbf{F}_1 = \mathbf{F}_4 = \mathbf{0}$. The initial quintiles are thought of as representing separate income classes and cease to be quintiles as they grow at different rates. The quintile-specific mean incomes and inequality measures in Table 1 are used to calculate overall inequality for each age group given the age-income distribution generated for each period. The values for 25–30-year-olds are assigned to the 15–30-year age group and the values for 40–45-year-olds are assigned to the 30–45-year age group. The income and fertility of the oldest men in each 15-year age interval are used in order to capture the largest proportion of child-bearing experience and to remove relationships between fertility and income due to age variation within the age group. For the survival rates in (12), average age-specific survival rates for Brazil are taken from the U.N. *Demographic Yearbook 1980* and are adjusted to introduce higher mortality for lower-income groups. Although survival rates in principle affect the distribution of income both within and across age groups, even large variations across income classes have negligible effects on the results reported here.

The most difficult data to obtain for even a simplified version of (12) are probabilities of sons' transitions across income classes conditional on their fathers' income class. Jose Pastore (1982) has estimated intergenerational mobility matrices for Brazil based on fathers' and sons' occupational categories. It is difficult to apply the results to income quintiles, however, since the sizes of the occupational categories vary from 2 to 65 percent of fathers, and since there is little guidance for converting Pastore's six occupational categories into income levels. Pastore's results nonetheless provide some guidance for constructing an empirically realistic mobility matrix.[22] Given the estimated fertility, survival, and mobility matrices for Brazil, the implied steady-state age-income distribution is generated by applying (12) to an initial distribution. The steady-state distribution implied by current fertility patterns is then used as the initial distribution for two counterfactual simulations, a 100 percent increase in the fertility of the bottom quintile, and a reduction in all quintiles' fertility rates to the level of the top quintile. The log variance and coefficient of variation for each age group are derived from the age-specific density functions for each period, using the within-quintile means and inequality measures reported in Table 1.[23]

---

[20] The total number of births are not used in the counterfactuals, but are included as a simple measure of the extent of differential fertility. The number of births actually applies to the current wife of the male household head, rather than to the male directly. It can be thought of as a measure of male-completed fertility, although it ignores divorce and the possibility of additional births after age 45. These errors should be small, however, and it is noteworthy that the 4.9 average births to male household heads aged 40–45 reported in Table 1 is identical to published estimates of Brazil's total fertility rate for 1977 (World Bank, 1979).

[21] Within-group inequality can in principle affect the results of the analysis, since, as shown in Section I, the effect of disproportionate growth of one class will depend in part on the within-group inequality for that class.

[22] Rather than allow mobility at every age, mobility is assumed to occur only when children become adults. This implies that every age-specific mobility submatrix $\mathbf{M}_i$ in (12) is an identity matrix except for $\mathbf{M}_2$, which is assumed to be

$$\begin{bmatrix} .50 & .25 & .15 & .10 & .05 \\ .25 & .40 & .20 & .20 & .10 \\ .15 & .20 & .35 & .20 & .20 \\ .05 & .10 & .20 & .35 & .25 \\ .05 & .05 & .10 & .15 & .40 \end{bmatrix}.$$

The diagonal elements dominate for every row and column, implying a relatively low degree of mobility, roughly consistent with Pastore's occupational mobility data. Many alternative forms for the mobility matrix were used for the counterfactuals, with results qualitatively similar to those presented below generated by any mobility matrix which is highly diagonal in structure. If high mobility is assumed, in which some classes are more likely to produce poor offspring than are the poor themselves, the results can change dramatically, since higher fertility for the poor can then decrease the proportion poor in transition periods and in the steady state.

[23] Another test is to take initial conditions which approximate the Brazilian age-income distribution in 1976 and compare the transition and steady state implied by the rates in Table 1 with the transition and steady state implied by an immediate change to equal fertility for all classes. The results demonstrate the same patterns as those in Table 2, with the log variance higher in all periods under equal fertility, and the coefficient of variation higher for all periods under the empirically observed rates. Note in Table 1 that the poorest quintile has a log mean more than one standard deviation from the overall log mean in both age groups. From the results in Section I, this implies that an expansion of the bottom quintile should initially increase log variance. The reason a change to equal fertility does not reduce the log variance, even initially, is that it also implies a reduction in the relative size of the second and third quintiles, both of which have log

TABLE 2—COUNTERFACTUAL SIMULATIONS OF EFFECTS OF DIFFERENTIAL FERTILITY ON INCOME INEQUALITY PREDICTED INEQUALITY MEASURES FOR BRAZILIAN MALE HOUSEHOLD HEADS AGED 30–45

	Case 1: Increase in Fertility of Poor			Case 2: Elimination of Fertility Differentials		
Period (1)	Percent Poor (2)	Coefficient of Variation (3)	Log Variance (4)	Percent Poor (5)	Coefficient of Variation (6)	Log Variance (7)
0	25.5	2.266	0.796	25.5	2.266	0.796
1	25.5	2.267	0.796	25.5	2.267	0.796
2	25.5	2.267	0.795	25.5	2.267	0.795
3	30.6	2.339	0.785	23.7	2.238	0.831
4	30.6	2.339	0.785	23.7	2.238	0.831
5	32.2	2.363	0.775	23.3	2.231	0.837
6	32.8	2.373	0.769	22.8	2.222	0.845
7	33.3	2.379	0.765	22.8	2.221	0.845
8	33.7	2.386	0.761	22.5	2.217	0.848
9	33.8	2.388	0.759	22.4	2.214	0.850
10	34.1	2.391	0.757	22.3	2.213	0.851
11	34.2	2.393	0.755	22.2	2.211	0.852
12	34.2	2.394	0.754	22.2	2.210	0.853
13	34.3	2.395	0.754	22.1	2.209	0.853
14	34.3	2.395	0.753	22.1	2.209	0.853
15	34.4	2.396	0.753	22.1	2.208	0.854

Notes: Poor refers to the class which was initially the poorest quintile in Table 1. The initial conditions for the counterfactuals are the steady state values implied by the fertility rates in Table 1. Case 1 assumes the fertility rates shown in Table 1 except for a 100 percent increase in fertility of the poorest class for both age groups. Case 2 assumes that the fertility rates of all classes are reduced to the rate for richest quintile in Table 1.

The results of these counterfactuals are shown in Table 2, with results reported only for the 30–45-year-old age group.[24] The values for period 0 show the percent poor, the coefficient of variation, and the log variance in the steady state implied by the income-specific fertility rates in Table 1 and the assumed mobility matrix and survival rates. Column 2 indicates that a 100 percent increase in the fertility of the poor leads to an increase in the percent poor, converging to a new steady-state value of 34 percent. This increase must occur, as shown in Section III, since the poor are more likely than any other class to produce poor offspring. Similarly, it is a necessary result, given the assumed mobility matrix, that the equal fertility regime of Case 2 produces a decrease in the percent poor from the initial value, as shown in column 5, and produces a smaller percent poor in every period than the high fertility regime of Case 1.

In Table 2, columns 3 and 6 show the coefficient of variation for the two counterfactuals. The coefficient of variation increases monotonically in response to the increase in fertility, as shown in column 3, while it declines monotonically in response to the reduction to equal fertility, shown in column 6. The results of the counterfactuals for the coefficient of variation, then, are consistent with the widespread notion that higher fertility for the poor will increase income inequality. The results for the variance of log income give just the opposite conclusion, however. Column 4 shows that the log variance declines monotonically to a new steady-

---

means within one standard deviation of the overall log mean. Although a decrease in the relative size of the poorest quintile alone would initially decrease the log variance, the net effect of equalizing fertility for all quintiles is to increase log variance, even in the short run.

[24] Recall that the 30–45-year-old age group takes on the income, inequality, and fertility levels of 40–45-year-olds, so that the results can be thought of as age-adjusted measures for 30–45-year-olds. Results for the 15–30-year-old age group demonstrate exactly the same qualitative patterns as those for 30–45-year-olds.

state value in response to the 100 percent increase in the fertility of the poor. Column 7, on the other hand, shows that the reduction to equal fertility causes a monotonic increase in log variance.

Although it is a surprising result, the fact that the coefficient of variation and the log variance move in opposite directions in response to changes in the fertility of the poor is consistent with the predictions of Sections I and II.[25] The results demonstrate the confusing signals given by inequality measures in the presence of income differentials in fertility. In the counterfactuals shown in Table 2, there is no change in the relative earnings of the poor relative to other classes. Changes in inequality measures occur only because of the changing population composition caused by the interaction of differential fertility and intergenerational mobility. If changes in relative earnings or the extent of intergenerational mobility are introduced into the model, it is easy to generate changes in inequality that are in the opposite direction of real welfare changes. Consider, for example, the hypothesis that higher fertility of the poor reduces the probability that poor children move into higher income classes. The reduction in log variance caused by an increase in the fertility of the poor in Table 2 can continue to appear even if the transition rates out of poverty (or the mean income of the poor) are reduced at the same time that the fertility of the poor is increased. The empirical results from Brazil thus confirm the theoretical predictions that inequality measures may give conflicting signals in the presence of differential fertility, and that those signals may be inconsistent with actual changes in relative welfare.

V. Conclusions

When population composition is constant, regressive income transfers produce unambiguous changes in standard measures of inequality. Intertemporal changes in relative welfare can therefore be inferred from intertemporal changes in inequality measures. The results above demonstrate that this correspondence between inequality indexes and relative welfare no longer holds when population composition is changing due to income differentials in fertility. The analytical results in Section I demonstrate the sensitivity of inequality measures to changes in the relative sizes of income classes, even when relative incomes are constant. The direction of the effect on inequality of a higher fertility rate for the poor may change over time and may be different for different inequality measures. The effects of population expansion on Lorenz curves, analyzed in Section II, indicate that intersecting Lorenz curves—and therefore ambiguous inequality comparisons—will be the rule rather than the exception whenever the introduction of new individuals into a population changes mean income.

Empirical evidence from Brazil indicates a regular pattern of differential fertility, with fertility falling monotonically as income rises across quintiles. The multistate life table constructed from these data to project the Brazilian male income distribution suggests that the empirical relationship between differential fertility and inequality is in fact as ambiguous as the theoretical discussion suggests. Counterfactual simulations comparing the effects of a 100 percent increase in the fertility of the poor with equal fertility across income classes indicate that higher fertility causes an increase in the coefficient of variation, but a decrease in the variance of log income, in both the steady state and in the transition.

These results raise serious concerns about the meaning of inequality comparisons when fertility varies systematically by income. Interpreting empirical relationships between population growth and inequality is especially problematic. The results imply, for example, that we cannot reject the hypothesis that population growth increases wages relative to rents, even if we observe an increase in the coefficient of variation whenever population growth rates increase. Standard inequality measures alone are inadequate to test hypotheses about the distributional effects of migration or high fertility, since it is impossible to distinguish the resulting changes in population composition from changes in relative welfare.

---

[25] At the initial conditions used for the counterfactuals in Table 2 the log mean for 30–45-year-olds in the poorest class is within one standard deviation of the overall log mean for the age group. Increases in the relative size of the class therefore decrease the log variance. The within-class mean is sufficiently far from the overall mean, however, that given the conditions on the coefficient of variation shown in (3), increases in the relative size of the poorest class increase the coefficient of variation. Although continued increases in the relative size of the poorest class would eventually cause the coefficient of variation to decline, as shown in Section I, the existence of intergenerational mobility insures convergence to a steady-state percent poor, which in this case occurs before the coefficient of variation reaches the turning point implied by (3).

The empirical application of the multistate life table for Brazil suggests that the direction of the compositional effects of differential fertility can in principle be identified. The data requirements, however, are extensive. An accurate test must account for intergenerational mobility, an issue for which there are few data, even in industrialized countries. While the empirical evidence from Brazil is far from complete, the results confirm the warnings of the theoretical analysis. The fact that the Brazilian data produce contradictory results for two standard inequality measures makes it clear that caution must be exercised in any empirical analysis of the distributional consequences of high fertility, with the compositional effects of differential fertility carefully accounted for.

REFERENCES

Adelman, Irma and Morris, Cynthia Taft, *Economic Growth and Social Equity in Developing Countries*, Stanford: Stanford University Press, 1973.

Ahluwalia, M. S., "Inequality, Poverty and Development," *Journal of Development Economics*, December 1976, *3*, 307–42.

Atkinson, Anthony B., "On the Measurement of Inequality," *Journal of Economic Theory*, September 1970, *2*, 244–63.

Boulier, Bryan L., "Population Policy and Income Distribution," in C. R. Frank and R. C. Webb, eds., *Income Distribution and Growth in the Less-Developed Countries*, Washington: The Brookings Institution, 1977.

Chu, Chin-Yi, "An Intergenerational Model of Bequests, Fertility, Income Distribution, and Population Growth," unpublished manuscript, Department of Economics, University of Michigan, 1984.

Fei, C. H., Ranis, Gustav and Kuo, Shirley W. Y., "Growth and the Family Distribution of Income by Factor Components," *Quarterly Journal of Economics*, February 1978, *92*, 17–53.

Fields, Gary S., "A Welfare Economic Approach to Growth and Distribution in the Dual Economy," *Quarterly Journal of Economics*, August 1979, *93*, 327–53.

Kakwani, Nanak C., *Income Inequality and Poverty: Methods of Estimation and Policy Applications*, Oxford: Oxford University Press, 1980.

Keyfitz, Nathan, *Applied Mathematical Demography*, New York: Wiley & Sons, 1977.

Lam, David, "The Variance of Population Characteristics in Stable Populations, With Applications to the Distribution of Income," *Population Studies*, March 1984, *38*, 117–27.

______, (1986a) "Distribution Issues in the Relationship Between Population Growth and Economic Development," in D. Gale Johnson and Ronald D. Lee, eds., *Population Growth and Economic Development*, Madison: University of Wisconsin Press, 1986.

______, (1986b) "Lorenz Curves and Inequality Comparisons Under Changing Population Composition," unpublished manuscript, University of Michigan, June 1986.

Land, Kenneth and Rogers, Andrei, *Multidimensional Mathematical Demography*, New York: Academic Press, 1982.

Morley, Samuel, "The Effect of Changes in the Population on Several Measures of Income Distribution," *American Economic Review*, June 1981, *71*, 285–94.

Parlett, Beresford, "Ergodic Properties of Populations I: The One Sex Model," *Theoretical Population Biology*, August 1970, *1*, 191–207.

Pastore, Jose, *Inequality and Social Mobility in Brazil*, Madison: University of Wisconsin Press, 1982.

Preston, Samuel H., "Differential Fertility, Unwanted Fertility, and Racial Trends in Occupational Achievement," *American Sociological Review*, August 1974, *39*, 492–506.

Ram, Rati, "Population Increase, Economic Growth, Educational Inequality, and Income Distribution," *Journal of Development Economics*, April 1984, *14*, 419–28.

Repetto, Robert, *Economic Equality and Fertility in Developing Countries*, Baltimore: Johns Hopkins University Press, 1979.

Robinson, Sherman, "A Note on the U Hypothesis Relating Income Inequality and Economic Development," *American Economic Review*, June 1976, 437–40.

______ **and Dervis, Kemal,** "Income Distribution and Socioeconomic Mobility: A Framework for Analysis and Planning," *Journal of Development Studies*, July 1977, *13*, 347–64.

Rodgers, Gerry, "Population Growth, Inequality and Poverty," *International Labour Review*, July-August 1983, *122*, 443–60.

Stark, Oded and Yitzhaki, Shlomo, "Migration, Growth, Distribution, and Welfare," *Economics Letters*, 1982, *10*, 243–49.

Swamy, Subramanian, "Structural Changes and the Distribution of Income by Size: The Case of India," *Review of Income and Wealth*, June 1967, *13*, 155–74.

Winegarden, C. R., "A Simultaneous-Equations Model of Population Growth and Income Distribution," *Applied Economics*, December 1978, *10*, 319–30.

______, "Socioeconomic Equity and Fertility in Developing Countries: A Block-Recursive Model," *De Economist*, NR 4, 1980, *128*, 530–37.

United Nations, *Demographic Yearbook 1980*, New York, 1981.

World Bank, *World Development Report*, Washington, 1979.

Estimates of the Direct Effects of Comparable Worth Policy

By GEORGE JOHNSON AND GARY SOLON*

Civil rights legislation of the early 1960's made it illegal for an employer to pay men and women on different bases for the same work, or to discriminate against women in hiring, job assignment, or promotion. Two decades later, however, the ratio of women's to men's earnings has shown little upward movement.[1] Furthermore, major sex differences in occupational distribution persist with predominantly female jobs typically paying less than predominantly male jobs. For example, in 1983 the percentages female among truckdrivers and secretaries were respectively 3.2 and 99.0, with the truckdrivers receiving considerably higher wages. Similar comparisons can be made of engineers vs. librarians, professors of economics vs. professors of art history, and literally hundreds of other examples of primarily men's jobs (*MJ*s) vs. primarily women's jobs (*WJ*s).

Such comparisons have generated much discussion, in both the judicial and political arenas, concerning the desirability of establishing "comparable worth" (CW) procedures for adjusting wage rates.[2] The presumption underlying the CW movement is that the observed negative relationship between wage rates and femaleness of an occupation reflects an "undervaluation" by society of *WJ*s relative to *MJ*s. Accordingly, CW policy—enacted through legislation or through judicial interpretation of Title VII of the Civil Rights Act—would require that the sex-neutrality of an employer's wage structure across jobs be assessed by "job evaluation" procedures. Specifically, each job within an organization would be assigned points in each of several dimensions (such as skill requirements, responsibility, effort, and working conditions), and these scores would somehow be aggregated to an overall index of the job's worth.[3] If it appeared that the employer paid systematically lower wages in *WJ*s than in *MJ*s with similar total scores, the employer would be in violation of the law and would be required to increase wages in *WJ*s. In statistical terms, then, CW essentially would require an employer to eliminate a negative partial correlation between wages and femaleness of occupation that persisted after controlling for legitimate determinants of pay.[4]

The object of this paper is to measure and analyze pay differences between *WJ*s and *MJ*s and to discuss how they would be affected by adoption of a CW policy. Section I presents empirical evidence on the relationship between wage rates and femaleness of occupation; Section II discusses the implications of this evidence for the potential impact of CW; and Section III summarizes our findings.

---

*Department of Economics, University of Michigan, Ann Arbor, MI 48109-1220. This study was partially supported by a grant from the Alfred P. Sloan Foundation to the Institute of Public Policy Studies at Michigan. We are grateful for comments from the referees, and seminar participants at Michigan, Cornell, Michigan State, Princeton, and the National Bureau of Economic Research.

[1] See June O'Neill (1985) for example.

[2] See Bureau of National Affairs (1981) for a useful review of the comparable worth doctrine and its relation to existing legislation, particularly the Equal Pay Act and Title VII of the Civil Rights Act. Recent progress by the comparable worth movement is summarized in Walter Goodman (1984) and Robert Pear (1985).

[3] Of course, this job evaluation process is extremely problematical. A detailed discussion in Donald Treiman and Heidi Hartmann notes that "there are no definitive tests of the 'fairness' of the choice of compensable factors and the relative weights given to them," and concludes that the process is "inherently judgmental" (1981, p. 96).

[4] Some economists have discussed CW as if it would require that all wages be directly *determined* by job evaluation studies, but our impression of most existing and proposed CW plans is that they would use job evaluation procedures merely to assess the sex-neutrality of wage structures.

I. Empirical Evidence

The presumed stylized fact that motivates CW proposals is the negative relation, *ceteris paribus*, between an individual's hourly earnings (W, which we express logarithmically) and the fraction of workers in his or her occupation who are women (F). A simple representation of this relation is

$$W = \beta_s' Z + \gamma_s F + \varepsilon, \tag{1}$$

where Z is a vector of characteristics that influence earnings, β_s is the associated coefficient vector for sex S, and ε is a disturbance term. The parameter γ_s reflects the wage-femaleness relationship for sex S, and $\exp(\gamma_s)$ represents the *ceteris paribus* wage ratio between a virtually all-female occupation and a virtually all-male occupation. The average log wage for each sex is $\overline{W}_s = \beta_s'\overline{Z}_s + \gamma_s \overline{F}_s$, and the male-female difference in average log wages can be expressed as

$$\begin{aligned} D &= \overline{W}_m - \overline{W}_f \\ &= \beta_m' \overline{Z}_m - \beta_f' \overline{Z}_f + \gamma_m \overline{F}_m - \gamma_f \overline{F}_f, \end{aligned} \tag{2}$$

where the m and f subscripts denote male and female, respectively.

This statistical framework is illustrated in Figure 1, which assumes for now that $\gamma_m = \gamma_f < 0$. The figure illustrates two salient features of male-female pay differences: 1) the negative slope of the wage functions indicates that occupations with higher F tend to pay less (which accounts for $W_* - \overline{W}_f$ of D in the figure), and 2) the separation between the male and female functions indicates that, given an occupation and its associated F, men in that occupation tend to earn more than women (which accounts for the remaining $\overline{W}_m - W_*$).

The latter phenomenon could be due to three possible factors. First, compared to women in the same occupation, men might have higher-earning average characteristics, such as longer average work experience. Second, some employers may pay female employees systematically less than male employees in the same occupation even when their average characteristics are the same. Third, the men in an occupation may tend to work for higher-paying firms while the women work for lower-paying firms. The first factor does not reflect discriminatory employment practices, but the second and third may. Indeed, the second was prohibited by the Equal Pay Act and has been illegal for over twenty years. The third may be at least partly related to the sorts of discriminatory hiring practices prohibited by the Civil Rights Act.

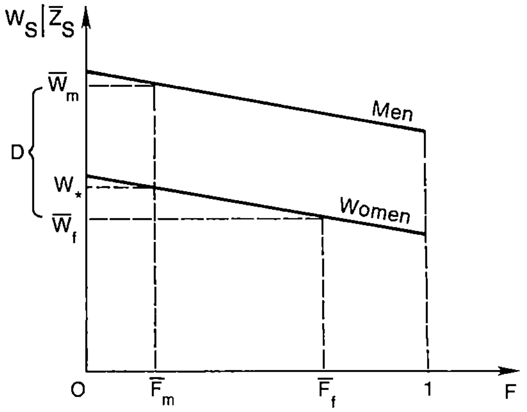


FIGURE 1

Despite the large quantity of research on male-female wage differences,[5] surprisingly few studies have attempted to assess the relative contributions of the second and third factors. What little evidence exists suggests that the third factor is far more important. Francine Blau's (1977) study of several narrowly defined clerical and computer-related occupations measured substantial male-female wage disparities in most of the occupations, but found they were due predominantly to between-firm wage differences, rather than to unequal pay within firms. Similar results were reported by John Buckley (1971) and Donald McNulty (1967).[6] If the small contribution of within-firm pay disparities is characteristic of most occupations, this may explain the seemingly negligible impact of the Equal Pay Act on the overall male-female wage gap.

---

[5] See, for example, Ronald Oaxaca (1973), Jacob Mincer and Solomon Polachek (1974), and Mary Corcoran and Greg Duncan (1979).

[6] Further discussion and references on this point can be found in Treiman and Hartmann (pp. 39–40).

What distinguishes comparable worth policy from existing antidiscrimination legislation is its goal of equalizing pay *across* occupations. In the context of Figure 1, one can view CW as an attempt to reduce the negative relationship between wages and femaleness of occupation, that is, to flatten the depicted wage functions. Although the *sources* of this relationship are not the main focus of this paper, they warrant some discussion. The two most prominent explanations for the downward slope of Figure 1's wage functions are sex differences in preferences and occupational exclusion.[7] According to the preferences explanation, workers have heterogeneous tastes for job characteristics, and the distribution of these tastes differs between men and women. These gender differences in job preferences might arise from socialization concerning sex roles and "appropriate" work (the nurse/doctor syndrome), differing family responsibilities, or differing expectations regarding continuous vs. intermittent labor market attachment. In any case, such taste differences would lead to sex differences in occupational distribution. Furthermore, if the preference distributions were such that the supply of workers to female-dominated occupations were especially large relative to the demand, these occupations would pay relatively low wages. In this analysis, the negative relationship between wages and femaleness of occupation results from voluntary choices, and the wage structure cannot be "improved" by legislative fiat. The relative wage changes mandated by a CW policy would be undesirable on efficiency grounds and not especially appealing on equity grounds.[8]

An alternative explanation for $\gamma < 0$ is that women are systematically excluded from many occupations and are thus "crowded" into a subset of occupations with depressed wage rates.[9] According to this explanation, women become secretaries, librarians, and art historians not by choice, but because they are blocked from becoming truckdrivers, engineers, and economists. The employer practices that would lead to such occupational exclusion are illegal under Title VII of the Civil Rights Act, and vigorous enforcement of that statute appears to be a more suitable long-run remedy than CW legislation. On the other hand, even the complete elimination of occupational exclusion would mainly benefit new female workers and might be of little help to older women already committed to traditionally female career paths. A CW policy might then be advocated on equity grounds as a means of compensating the latter group.

Whatever the *source* of the negative relationship between wages and femaleness of occupation, the remainder of this section attempts to *quantify* that relationship. Measuring the relationship will serve, first, to indicate the magnitude of the phenomenon that CW seeks to remedy and, second, to facilitate the next section's analysis of CW's potential impact. We therefore have estimated equation (1) with May 1978 *Current Population Survey* (*CPS*) data on nonagricultural wage and salary workers, at least 16-years-old, who responded to the supplementary questions concerning their "usual weekly earnings" and "usual weekly hours." Relative to other surveys, the *CPS* has the advantage of detailed occupational information on a large national sample. Relative to subsequent *CPS* samples that were asked the supplementary questions, the May 1978 sample has the advantages of a larger size and a more "normal" unemployment situation.

The dependent variable in our regression analyses is the natural logarithm of the ratio of usual weekly earnings to usual weekly hours. This variable, W, averages 1.757 for the 24,056 men in our sample and 1.346 for the 19,412 women. This implies that D in equation (2) is $1.757-1.346=.411$, which means that the (geometric) mean wage for women is 33.7 percent less than that for men. Our femaleness variable, F, is the proportion female in the individual's 3-digit occupation. Whenever possible, this variable was obtained from 1978 *CPS* data as reported in the January 1979 *Employment and Earnings* from the Bureau of Labor Statistics (BLS). In occupations with *CPS* cells so small that BLS deemed the F estimates too unreliable to publish, we extrapolated from 1970 census data. The resulting F variable ranges from less than .01 (for such categories as firefighters, automobile mechanics, and plumbers and pipefitters) to over .99 (secretaries and housekeepers).

---

[7] A third explanation is that wages for some *WJ*s, especially hospital nurses, are depressed because employers have monopsony power. There is little evidence, though, that this analysis applies to enough *WJ*s to provide a general explanation.

[8] More detailed discussions appear in Mark Killingsworth (1984) and Jennifer Roback (1986).

[9] See Barbara Bergmann (1971, 1974).

The other variables, comprising the Z vector in equation (1), are: years of school completed; potential work experience (age minus years of school minus six)[10] and its square; regional dummy variables for Northeast, North Central, and West; three dummies for residence in a large, medium, or small SMSA; race dummies for black or other minority race; dummies for voluntary and involuntary part-time work; two marital status dummies, one for married and the other for separated, widowed, or divorced; number of children and a dummy for presence of children; dummies for union membership and, if not a member, coverage by a union contract; a government employment dummy; dummies for 20 major industries (mining, construction, durable goods manufacturing, etc., with private households as the omitted category); the fraction of workers in the individual's occupation who worked less than 30 hours in the 1970 census week; and a series of occupational characteristics indices, developed by the staff of a National Research Council committee,[11] describing the "general educational development," "specific vocational preparation," strength, physical demands, and undesirable environmental conditions associated with the occupation. The committee itself acknowledged that these latter variables are arbitrary and of dubious validity; on the other hand, they are qualitatively similar to the job evaluation scores that would be applied under a CW policy and therefore ought to be controlled for in the estimation of the γ's.

Indeed, the general question of whether certain variables should be included in the Z vector comes down to whether CW would allow employers to base pay on those variables. If so, attention would focus on the *partial* relationship between F and wages while controlling for those variables. Consequently, the industry dummies undoubtedly belong in the regressions because CW would require equal pay for work of comparable worth only *within* firms. Wage differences *between* firms would not be covered, and a fortiori industry wage differences would be unaffected. On the other hand, whether union status belongs in Z depends on whether CW would permit an employer to pay union workers in one occupation on a different basis than nonunion workers in another occupation. Presumably, variables such as race and marital status do not belong in our regressions except as proxies for other characteristics that employers would be allowed to use. Given the considerable ambiguity in the choice of Z variables, we will make careful efforts in the analysis below to clarify how different control variables affect the estimation of the γ's.

We begin by reporting estimates, for each sex, of simple regressions of W on F. These estimates describe the *gross* relationship between wages and femaleness of occupation. For men, the estimated coefficient for F, $\tilde{\gamma}_m$, is $-.343$ (with estimated standard error .015), which implies a .710 wage ratio between virtually all-female and all-male occupations. For women, the estimated coefficient $\tilde{\gamma}_f$ is $-.244$ (.013), implying a .783 wage ratio. These results confirm Donald Treiman and Heidi Hartmann's (1981) findings from aggregate data that earnings are negatively related to femaleness of occupation for both men and women, and that the relationship is stronger among men. (The latter finding is discussed further in the next section.)

---

[10] The unavailability of an actual experience measure is a major drawback of our data set. Daniel Hamermesh has suggested to us that women who have spent more time out of the labor force may be more concentrated in occupations with high F. If so, our estimates below are biased toward a larger negative coefficient of F in the women's wage equation and therefore tend to overstate the size of CW's target.

[11] Ann Miller et al. (1980, appendix F).

These estimates, however, take no account of differences in workers' characteristics or the characteristics of their jobs. Therefore, we next report the main results from estimating multiple regressions of W on F controlling for other variables.[12] To begin with, controlling for the Z vector described above reduces the estimated F coefficients by more than one-half. For men, controlling for Z reduces the estimated coefficient of F from $\tilde{\gamma}_m = -.343$ to $\hat{\gamma}_m = -.168$ (.015); for women, the reduction is from $\tilde{\gamma}_f = -.244$ to $\hat{\gamma}_f = -.090$ (.014). The new coefficients imply *ceteris paribus* wage ratios between virtually all-female and all-male occupations of .845 for men and .914 for women.

To clarify how the inclusion of various control variables affects the estimation of the γ's, we use the fact that the change from a simple to a multiple regression estimate of γ can be expressed as

$$\hat{\gamma} - \tilde{\gamma} = -\sum_{j=1}^{K} \hat{\beta}_j b_{jF}, \tag{3}$$

where the $\hat{\beta}_j$'s are the estimated coefficients of the K control variables in the log wage regression and the b_{jF}'s are the coefficients from auxiliary simple regressions of the control variables on F.[13] Equation (3) enables a straightforward decomposition of the difference $\hat{\gamma} - \tilde{\gamma}$ into the parts attributable to each control variable. Table 1 summarizes the decompositions for $\hat{\gamma}_m - \tilde{\gamma}_m$ and $\hat{\gamma}_f - \tilde{\gamma}_f$. The striking finding is that, for both men and women, the bulk of the reduction from $\tilde{\gamma}$ to $\hat{\gamma}$ is due to the industry dummy variables. A closer look at the underlying data reveals the main reason for this industry effect: workers in construction and manufacturing (especially durables) are relatively well paid given their other characteristics and are predominantly male. Once the wage effect of belonging to these industries is separately accounted for, the remaining effect of femaleness of occupation is considerably reduced.

TABLE 1—DECOMPOSITION OF THE INFLUENCE OF CONTROL VARIABLES ON THE ESTIMATION OF γ

	Men	Women
Total Effect = $\hat{\gamma} - \tilde{\gamma}$	.175	.153
Decomposition		
Schooling and Experience	−.024	.024
Region	−.002	−.001
SMSA Variables	−.015	−.001
Race	.003	.001
Part-Time Status	.019	.009
Marital Status	.031	.002
Children	.003	.001
Union Status	.029	.022
Government Employment	.015	.001
Industry	.152	.083
Occupational Characteristics	−.034	.012

It is crucial to understand that, even if the relation between wages and femaleness of *industry* arises from discriminatory exclusion practices, CW would not remedy the resulting pay differences. As discussed above, CW would require equal pay for work of comparable value only within firms. Interindustry wage differences would be unaffected.[14] Consequently, the smaller γ estimates from the multiple regression analyses are clearly better indicators of the size of CW's target. If anything, they probably exaggerate CW's target because they control only for industry effects and not firm effects. An ideal data base that enabled controls for firm effects as well could also take account of CW's inability to alter pay differences between firms in the same industry.[15]

---

[12] The full results are reported in our earlier paper (1986).

[13] See Arthur Goldberger (1964, pp. 194–95).

[14] Robert Gregory and Ronald Duncan (1981) discuss an Australian policy akin to CW and find that it has had major relative wage effects. Unlike CW proposals in the United States, however, this policy does apply across firms.

[15] Indeed, if CW legislation were enacted, its inapplicability to interfirm differences might be exploited through business reorganizations. For example, a firm might "contract out" its female-dominated clerical functions to another firm to preclude pay comparisons with its other job categories.

It can certainly be argued that some of the variables included in the multiple regressions should be excluded. For example, a stringent CW law might require firms to pay their nonunionized employees comparably with their unionized employees. If so, the unionization variables should be excluded from the regressions. Similarly, although some variables such as those for marital status may proxy for legitimate determinants of pay, they may not belong in the regressions in their own right.

Therefore, to check the robustness of our γ estimates, we have estimated a parsimonious model that controls only for schooling, experience, region, SMSA, government employment, industry, and occupational characteristics other than fraction part-time. The resulting γ estimates, $-.176$ (.015) for men and $-.085$ (.014) for women, are very similar to those in the previous model. Also, to account for the outside possibility that pay under CW could be based only on *job* characteristics and not on personal characteristics,[16] we have estimated an even more parsimonious model that excludes the schooling and experience variables. With this specification, the estimated γ's are $-.233$ (.016) for men and $-.093$ (.014) for women.

Finally, we have conjectured that controlling for firm effects would lead to smaller estimates of CW's target. As a step in that direction, we have reestimated the full model with a more detailed set of 48 industry dummy variables. As expected, the estimated γ's decline in magnitude to $-.160$ (.015) for men and $-.068$ (.014) for women. Adding the more detailed industry dummies to the parsimonious model results in estimated γ's of $-.157$ (.015) for men and $-.057$ (.014) for women.

Some qualitative implications of these results are immediately clear. In every specification, a negative relationship between wages and femaleness of occupation appears, in accordance with the concerns of the CW movement. The reduction in the relationship's magnitude when industry is controlled for, however, highlights the point that, since CW would not apply across industries (or, indeed, across firms), the portion of the wage-femaleness correlation that is remediable by CW is smaller than might otherwise be supposed. A fuller discussion of implications regarding the potential impact of CW is given in the next section.

II. Implications for Comparable Worth Policy

Our empirical analyses, as well as those by other researchers, have established several important regularities in male-female wage differences. First, wages are negatively related to femaleness of occupation. Second, as indicated in all the results in the previous section, this relationship is stronger among men than among women. Third, as found by the Blau and other studies cited above, male-female pay differences *within* occupations are predominantly interfirm, not intrafirm.

This last point is crucial for understanding the potential impact of CW. If male-female pay differences within occupations are interfirm, a CW policy that applies only to intrafirm pay disparities can do little about the discrepancy between the male and female wage functions in Figure 1. Instead, CW's purview would be limited essentially to flattening the slopes of the two functions. Indeed, if the figure's assumption that $\gamma_m = \gamma_f = \gamma$ were correct, CW's direct effect on D in equation (2) could be well approximated by $\gamma(\bar{F}_f - \bar{F}_m)$.

Estimating CW's direct impact is complicated, however, by the robust finding that $\gamma_m < \gamma_f$. This raises an issue reminiscent of Ronald Oaxaca's (1973) problem of whether to use men's or women's wage equation coefficients in assessing how much of the male-female wage gap reflects sex differences in characteristics. Oaxaca's solution of using both to establish a range of estimates has some appeal here, too. In fact, in our earlier paper, we have developed a formal model of wage structure in the context of which we prove that the true direct impact of CW on D is bracketed by $\gamma_m(\bar{F}_f - \bar{F}_m)$ and $\gamma_f(\bar{F}_f - \bar{F}_m)$.[17] If we then apply the sample values $\bar{F}_f = .71$ and $\bar{F}_m = .21$ and the estimates $\hat{\gamma}_m = -.160$ and $\hat{\gamma}_f = -.068$ from the multiple regressions with detailed industry controls, the estimated direct effect of an economy-wide CW program is to reduce the existing D of .411 by between .034 and .080. Whether this initial effect strikes one as large or small depends, of course, on one's prior expectations of CW's impact.[18]

---

[16] The National Research Council report on comparable worth notes that "job evaluation plans currently in use do not ordinarily include worker characteristics because these plans typically attempt to measure only the required elements of jobs.... If job evaluation procedures are to be used for the resolution of claims of pay discrimination, however, their usefulness would be enhanced by including those worker characteristics regarded as legitimate compensable factors" (see Treiman and Hartmann, p. 86).

These estimates, however, overlook the likelihood that even a relatively comprehensive CW law would, like existing antidiscrimination legislation, be effectively confined to the public sector and large private firms. Such partial coverage would attenuate CW's direct effects. If CW covered a fraction κ of the jobs in the economy and if (for lack of an obviously superior assumption) an employer's coverage status were independent of its previous wage/femaleness relationship, CW's direct wage effect would be simply κ times the effect estimated above. For example, if $\kappa = .4$ (slightly above the figure suggested by Walter Oi, 1985), CW's estimated direct effect would be to reduce D by between .014 and .032. In other words, whereas the geometric mean wage for women previously was 33.7 percent less than that for men, after CW this shortfall would lie between 31.5 and 32.8 percent. While one's position on CW presumably would still depend on issues of principle, it is striking that CW's practical effect on the male-female wage gap would be so small.

We believe our characterization of CW's direct impact is a sensible approximation, but it must be qualified by several important caveats. First, we have not addressed such issues as how effectively CW legislation would be enforced or whether it would influence compliance with other antidiscrimination laws. Given the inherent ambiguities of the job evaluation process, these implementation issues could be of considerable practical importance.

Second, we have assumed that CW would attack the partial relationship between wages and femaleness of occupation, but would leave unaltered the dependence of wages on legitimate personal and job characteristics. It is conceivable that whatever job evaluation procedures gain favor in the course of case-by-case litigation would produce fundamental changes in this dependence, but these changes are virtually impossible to foresee.

Third, our estimates pertain only to *direct* effects. In competitive markets, however, CW-imposed wage increases in *WJ*s would lead subsequently to wage changes in *MJ*s, higher output prices, and substitution of employment in *MJ*s for employment in *WJ*s. Once account is taken of these general equilibrium effects (which are analyzed in detail in our earlier paper), holders of covered *WJ*s would enjoy relative wage increases, but workers in uncovered *WJ*s (including workers displaced by employment adjustments in the covered sector) would suffer relative wage decreases.[19] In fact, with sufficiently elastic labor demand, the latter effect could dominate and cause a decrease in overall average wages in *WJ*s relative to those in *MJ*s.[20]

Finally, lest there be any misunderstanding, we want to emphasize that our results concerning CW's direct impact say nothing about the extent of sex discrimination in the labor market. Our estimates of the relationship between wages and femaleness of occupation, though useful for inferring CW's potential impact, overlook some forms of sex discrimination and also may encompass some nondiscriminatory factors. For example, sex discrimination in hiring by high-paying employers may contribute to the male-female wage gap, but it is beyond the reach of CW policy (though it is subject to Title VII of the Civil Rights Act). On the other hand, as discussed in Section I, the negative wage/femaleness relationship we have measured may arise partly from sex differences in preferences rather than from discriminatory employment practices.

---

[17] This model uses wage/turnover tradeoffs to generate an equilibrium account of all three empirical regularities cited at the beginning of this sector. It can be viewed as a formalization of ideas in Blau (especially the last paragraph on p. 34).

[18] These estimates are somewhat smaller than those of Elaine Sorenson (1986), who considered potential wage adjustments in four state governments that conducted CW-related job evaluation studies. One reason for the discrepancy may be simply that CW's effects would be larger in the public sector than in the private sector. Our earlier paper contains some supporting evidence. Furthermore, it stands to reason that those state governments in the forefront of the CW movement would perform job evaluation studies conducive to especially large wage adjustments in *WJ*s.

[19] These results are similar to those in the literature on minimum wages. See Finis Welch (1976) for example.

[20] Labor demand in covered *WJ*s might be particularly elastic because of the possibility of "contracting out," as described in fn. 15.

III. Conclusions

This study has found that, when relevant worker and job characteristics are controlled for, a negative relationship between wages and femaleness of occupation still appears, in accordance with the concerns of the comparable worth movement. The impact of CW on male-female pay differences, however, would be limited by the policy's inapplicability to pay disparities between firms. Estimates that take account of this limitation suggest that a CW policy with partial coverage would eliminate only a small fraction of the gap between average male and female wages.

REFERENCES

Bergmann, Barbara R., "The Effects on White Incomes of Discrimination in Employment," *Journal of Political Economy*, March/April 1971, *79*, 294–313.

______, "Occupational Segregation, Wages and Profits When Employers Discriminate by Race or Sex," *Eastern Economic Journal*, April/July 1974, *1*, 103–10.

Blau, Francine D., *Equal Pay in the Office*, Lexington: D.C. Heath, 1977.

Buckley, John E., "Pay Differences between Men and Women in the Same Job," *Monthly Labor Review*, November 1971, *94*, 36–39.

Corcoran, Mary and Duncan, Greg J., "Work History, Labor Force Attachment, and Earnings Differences between the Races and Sexes," *Journal of Human Resources*, Winter 1979, *14*, 3–20.

Ehrenberg, Ronald G. and Smith, Robert S., "Comparable Worth in the Public Sector," NBER Working Paper No. 1471, September 1984.

Goldberger, Arthur S., *Econometric Theory*, New York: Wiley & Sons, 1964.

Goodman, Walter, "Equal Pay for 'Comparable Worth' Growing as Job-Discrimination Issue," *New York Times*, September 4, 1984, p. 17.

Gregory, Robert G. and Duncan, Ronald C., "The Relevance of Segmented Labor Market Theories: The Australian Experience of the Achievement of Equal Pay for Women," *Journal of Post-Keynesian Economics*, Spring 1981, *3*, 403–28.

Johnson, George and Solon, Gary, "Pay Differences between Women's and Men's Jobs: The Empirical Foundations of Comparable Worth Policy," Discussion Paper No. 254, Institute of Public Policy Studies, University of Michigan, August 1986.

Killingsworth, Mark R., "Heterogeneous Preferences, Compensating Wage Differentials and Comparable Worth," unpublished, June 1984.

McNulty, Donald J., "Differences in Pay between Men and Women Workers," *Monthly Labor Review*, December 1967, *90*, 40–43.

Miller, Ann R. et al., *Work, Jobs, and Occupations: A Critical Review of the Dictionary of Occupational Titles*, Committee on Occupational Classification and Analysis, NRC, Washington: National Academy Press, 1980.

Mincer, Jacob and Polachek, Solomon, "Family Investments in Human Capital: Earnings of Women," *Journal of Political Economy*, Suppl., March/April 1974, *82*, S76–108.

Oaxaca, Ronald, "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, October 1973, *14*, 693–709.

Oi, Walter Y., "Costs and Consequences of Comparable Worth," unpublished, May 1985.

O'Neill, June, "The Trend in the Male-Female Wage Gap in the United States," *Journal of Labor Economics*, Suppl., January 1985, *3*, S91–116.

Pear, Robert, "Women Vow to Continue Battle for Pay Equity after Legal Loss," *New York Times*, September 6, 1985, p. 7.

Roback, Jennifer, *A Matter of Choice: A Critique of Comparable Worth by a Skeptical Feminist*, New York: Priority Press, 1986.

Sorenson, Elaine, "Implementing Comparable Worth: A Survey of Recent Job Evaluation Studies," *American Economic Review Proceedings*, May 1986, *76*, 364–67.

Treiman, Donald J. and Hartmann, Heidi I., *Women, Work, and Wages: Equal Pay for Jobs of Equal Value*, Committee on Occupational Classification and Analysis, NRC, Washington: National Academy Press, 1981.

Welch, Finis, "Minimum-Wage Legislation in the United States," in Orley Ashenfelter and James Blum, eds., *Evaluating the Labor-Market Effects of Social Programs*, Princeton: Industrial Relations Section, Princeton University, 1976, 1–38.

Bureau of National Affairs, Inc., *The Comparable Worth Issue*, Washington, 1981.

The Role of Foreign Direct Investment in U.S. Capital Outflows

By Roy J. Ruffin and Farhad Rassekh*

United States foreign direct investment (FDI) is the sum of equity in and net claims on the foreign affiliates of U.S. firms. The affiliates are those entities over which the U.S. parent has direct or indirect ownership amounting to 10 percent or more. Portfolio investment consists of all other claims on foreigners. The purpose of this paper is to test the hypothesis that foreign direct investment and foreign portfolio investment are perfect substitutes. In other words, capital may be perfectly fungible.

The evaluation of the multinational corporation (MNC) frequently supposes that foreign direct investment does not displace any portfolio investment and simply adds to the existing stock of capital invested abroad (see Gary Hufbauer, 1975, pp. 290–91, or Richard Caves, 1971). But there are two aspects of a MNC: an international transfer of production facilities and a transfer of capital. It has often been observed, moreover, that the spread of MNCs need not involve any such transfer of capital—that a MNC can always seek financing in the host country (see Charles Kindleberger, 1969, and Neil Hood and Stephen Young, 1979). Therefore, the phrase "foreign direct investment" may be a misnomer (see Ruffin, 1984, p. 248), and the phenomena to which it refers may be a substitute for portfolio investment.

This paper estimates a simple model of U.S. portfolio capital outflows that builds on the work of William Branson (1968), James Tobin (1969), Pentti Kouri and Michael Porter (1974), and Peter Kenen (1978). The empirical results are consistent with the hypothesis that every dollar of U.S. FDI results in one less dollar being invested in foreign portfolio investment. Thus, the way MNCs finance their operations may be irrelevant to the net flow of capital between countries.

Section I develops the basic hypothesis, Section II presents a modified version of the portfolio balance model of Kouri and Porter, and Section III presents the empirical results.

I. Hypotheses

The purpose of MNCs is to internalize the returns to the ownership-specific assets of the firm (Hood and Young; Peter Buckley and Mark Casson, 1976). For example, the firm may possess certain marketing skills (Sears), hold certain trade secrets (Coca-Cola), or own certain patents (IBM) that enable it to compete effectively in distant foreign markets. Because information and monitoring costs are significant, MNCs find it more profitable to locate abroad than to simply license foreign firms or to rent out management consulting teams.

According to the internalization hypothesis, the establishment of an MNC is explained by numerous microeconomic factors (mostly unobserved) dealing with the economics of industrial organization. Similarly, the financial decisions of MNCs will involve a host of microeconomic variables involving the details of corporation finance. Explanations of these phenomena would have to be cross sectional in nature, drawing on the characteristics of each company. Indeed, it has been difficult to find any correlation of interest rate differentials with FDI in a time-series analysis (see Kenen and Hufbauer). If FDI is a function of many unobserved microeconomic variables, which are uncorrelated with the relevant macroeconomic variables, aggregate FDI will be exogenous in a time-series analysis of the portfolio balance model.

---

*Departments of Economics, University of Houston, Houston, TX 77004 and University of St. Thomas, Houston, TX 77006, respectively. We are indebted to the referee for many valuable comments. We also thank, without implicating, Badi Baltagi, Richard Caves, Pulak Das, Gerald Dwyer, Paul Evans, David Papell, and Khalil Yazdi. A data appendix is available on request.

If capital is fungible and not firm specific, when a MNC invests one dollar abroad the home and foreign capital markets are temporarily disturbed as interest rates would rise at home or fall aroad: To restore the original interest rates, portfolio investment abroad would fall by exactly one dollar.

II. A Model of Capital Outflows

To test the hypothesis that FDI is a perfect substitute for portfolio investment, we need only include FDI as an exogenous variable in an equation that otherwise accounts for portfolio investment. In the portfolio balance model, real asset demands are functions of real interest rates, income, wealth, and risk factors. Equilibrium occurs when interest rates adjust to clear all asset markets. Capital flows can be estimated simply by considering the domestic demand for foreign securities and the foreign demand for domestic securities (see Branson; Kouri and Porter; and Maurice Obstfeld, 1983).

A. *Portfolio Balance*

In a world of uncertainty, home and foreign investments will be imperfect substitutes. Let r be the real rate of return to domestic investment; and let r^* denote the real rate of return domestic residents earn by investing in the foreign country. The real interest rate r is exactly defined by

$$r = (i - \pi)/(1 + \pi), \tag{1}$$

where i is the nominal interest rate and π is the expected inflation rate in the home country. The expected real interest rate earned by domestic investors in the foreign country is

$$r^* = (1 + i^*)E/(1 + \pi)S - 1, \tag{2}$$

where i^* is the appropriate foreign interest rate, E is the expected exchange rate on the maturity date of the foreign bond, and S is the spot exchange rate on the date of the investment. In order to capture the impact of default risk as well as exchange risk, i and i^* are defined as 90-day commercial paper rates or banker's acceptances, rather than as Treasury bill rates.

Let Z denote the real quantity of private assets (i.e., those assets held by U.S. private agents) held abroad; similarly, Z^* is the real quantity of private assets the foreign country owns in the home country; let W be the quantity of real financial wealth owned by the home country; and let y be home real income. The modified version of the portfolio balance model used here is

$$Z = F(\overset{-}{r}, \overset{+}{r^*}, \overset{?}{y}, \overset{+}{W}, \overset{+}{Z^*}). \tag{3}$$

The variable r^* is a vector and Z^* is the sum of investments made by the rest of the world in the home country.

The variable Z^* is added to reflect two institutional features of the international capital market. First, Eurodollar rates exceed U.S. deposit rates, which encourages capital outflows, and the U.S. prime rate exceeds the London interbank offered rate (Libor), which encourages capital inflows (Russell Krueger, 1981). Second, the United States as a world banker implies the exchange of short-term liabilities for long-term assets (Emil Despres, Kindleberger, and Walter Salant, 1966). Thus, institutional characteristics of the international capital market imply a strong link between capital imports and capital exports.

The condition for money market equilibrium in the home country implies that the domestic rate of return r will be influenced by y, as in Kouri and Porter. The demand for real money balances, L, depends on y and the nominal interest rate, i. Thus, money market equilibrium holds in the home country when

$$m = L(\overset{-}{i}, \overset{+}{y}) = L(\pi + r + r\pi, y), \tag{4}$$

where m is the (in the short run) exogenously given real money supply. Equations (1), (3), and (4) can explain Z as a function of π, the r^*, the m, the y, W, and Z^*. Thus:

$$Z = Z(\overset{+}{\pi}, \overset{+}{r^*}, \overset{+}{m}, \overset{?}{y}, \overset{+}{W}, \overset{+}{Z^*}). \tag{5}$$

The signs of the effects of each variable on the right-hand side are intuitively obvious. For example: an increase in the home country's m, *ceteris paribus*, affects capital flows by decreasing r, which induces agents to hold a larger quantity of foreign securities. An increase in π, holding y and m constant, must lower r (equation (4)) and increase Z.

We cannot, of course, estimate equation (5) as it stands. A simple solution is to linearize (5) by taking first differences, that is,

$$\text{(6)}\quad \Delta Z = (\partial Z/\partial \pi)\Delta\pi + (\partial Z/\partial m)\Delta m + (\partial Z/\partial r^*)\Delta r^* + (\partial Z/\partial y)\Delta y + (\partial Z/\partial W)\Delta W + (\partial Z/\partial Z^*)\Delta Z^*.$$

B. *Specific Hypotheses*

The final regression equation that we will estimate can now be stated. Substitute $\Delta Z = \Delta Z_p + \Delta Z_d$ into (6), where Z_d and Z_p are direct and portfolio investment, and put ΔZ_p on the left-hand side; thus:

$$\text{(7)}\quad \Delta Z_p = a_0 + a_1\Delta\pi + a_2\Delta m + a_3\Delta y + a_4\Delta r^* + a_5\Delta Z_d + a_6\Delta W + a_7\Delta Z^* + a_8\Delta G + a_9 D + u.$$

A constant term is added to capture the effects of omitting variables with a time trend; u is an error term.

Two *ad hoc* variables are added to equation (7): the real price of gold, G, serves as a proxy for speculation and the dummy variable D captures the impact of capital controls. Germany had controls on capital inflows until 1974:I and the United States had controls on capital outflows until 1974:I.

We expect the signs

$$a_1 > 0,\ a_2 > 0,\ a_3 < \text{ or } > 0,\ a_4 > 0,$$

$$a_5 = -1,\ a_6 > 0,\ a_7 > 0, a_8 < 0,\ a_9 < 0.$$

Of course, if $a_5 = -1$, foreign direct investment perfectly substitutes for portfolio investment.

III. Empirical Estimates

To estimate equation (7) we chose as our sample period the quarters 1970:III to 1983:I. We are interested in the behavior of U.S. private portfolio assets abroad and U.S. private foreign direct investment abroad. As a representative sample of foreign rates of return, we used the United Kingdom, Canada, and West Germany. Presumably, the behavior of real interest rates in these countries are indicative of the major foreign investment opportunities facing the United States.

We make the assumption that the United States is a small country, so that the elements of r^* are exogenous. Is this reasonable? Yes. The U.S. net capital outflow may only be about 1 or 2 percent of the total world capital market. For example, in 1980, the U.S. net capital outflow was only \$31 billion (the value of world trade in 1980 was \$2 trillion). Indeed, the simple correlation coefficients between first differences of the net capital outflow and real rates of return in the United Kingdom, Canada, and West Germany are nearly zero.

The variable ΔZ^* is endogenous, but it is difficult to specify a simple structural equation because so many countries are involved (see Kenen, p. 38). Thus, the endogenous components of ΔZ^* are purged by the method of instrumental variables.

To estimate (7) we require estimates of the U.S. inflation rate directly in the variable π and indirectly in the r^* variables (see equation (2)). We applied the technique of Box and Jenkins to estimating the expected inflation rate for the United States. An ARIMA model (Charles Nelson, 1973) was fitted to the U.S. inflation rate for the period 1957:I to 1983:II. The R^2 was 0.82. The expected inflation rate was defined as the four-step (one-year) ahead forecasted inflation rate from the fitted time-series model, which presumably gives us the best estimate of future inflation.

Equation (2) requires expected exchange rates for the U.S. dollar against the currencies of West Germany, Britain, and Canada. Fortunately, forward exchange rates can be used to approximate expected exchange rates (see, for example, Obstfeld). While this procedure is approximate (it assumes risk neutrality), it is superior to estimating expected exchange rates by strictly statistical techniques such as Box-Jenkins.

A. *Regression Results*

Table 1 shows the results of applying instrumental variables to equation (7) (see Ray Fair, 1970). While the coefficients generally have the expected signs, it is interesting to observe that the interest rate variables (including π) and financial wealth are insignificantly different from zero.

Money has a powerful impact on U.S. capital outflows. Each \$1 increase in U.S. *M*1 causes \$.06 in capital outflows to the rest of the world. This estimate is highly significant.

The coefficient of ΔZ_d casts light on the perfect substitutability hypothesis. As argued above, if MNCs do not influence total capital outflows, this coefficient should be approximately -1 when the dependent variable is the outflow of portfolio capital. The actual coefficient is -1.009 with a standard error of .32. Thus, we can reject the hypothesis that portfolio investment is independent of direct foreign investment; but we cannot reject the hypothesis that direct foreign investment substitutes, dollar for dollar, for portfolio investment.[1]

The coefficient of real U.S. capital inflows, ΔZ^*, is .7904 and highly significant, reflecting the mutual dependence between outflows and inflows arising from the institutional features of the international capital market. Thus, the theory of capital outflows that receives the most support is that such outflows reflect the role of the United States as a banker to the world as well as various interest rate controls or customs in different countries.

TABLE 1—REGRESSION COEFFICIENTS

	Coefficient	Standard Error	*t*-Ratio
Constant	3.6824	1.2323	2.9884
$\Delta\pi_{US}$	.112	.1778	.6303
Δr_{WG}	.446	1.1131	.4002
Δr_{UK}	1.319	1.0556	1.2498
Δr_{CA}	-1.825	1.4750	-1.2374
Δm_{US}	.0628	.0284	2.2176
Δy_{US}	$-.0395$	.025	-1.580
ΔZ_d	-1.0093	.3176	-3.1779
ΔW	.0127	.0104	1.2222
ΔZ^*	.7904	.1209	6.5387
ΔG	$-.0190$	.0118	-1.6088
D	-3.6396	2.0553	-1.7704

Note: Dependent Variable: ΔZ_p; Adj. $R^2 = .74$; Durbin-Watson = 2.04.

V. Conclusions

Within the framework of a modified portfolio-balance model, we have not been able to reject the hypothesis that foreign direct investment displaces an equal amount of portfolio investment. Thus, multinational corporations may have no significant impact on net capital flows. Moreover, the welfare impact of MNCs cannot be examined along the conventional lines of the costs and benefits of foreign investment.

The next step would be to examine the sensitivity of our conclusion to the use of different models of capital flows. The strong interaction between capital inflows and capital outflows suggests the necessity of an explicit model of the institutional structure of capital inflows and outflows.

REFERENCES

Branson, William H., *Financial Capital Flows in the U.S. Balance of Payments*, Amsterdam; North-Holland, 1968.

Buckley, Peter J. and Casson, Mark, *The Future of the Multinational Enterprise*, London: MacMillan, 1976.

Caves, Richard E., "International Corporations: The Industrial Economics of Foreign Investment," *Economica*, February 1971, *38*, 1–27.

Despres, Emil, Kindleberger, Charles P. and

---

[1] We also ran the regression with net portfolio outflows as the dependent variable and net FDI on the right-hand side, and, of course, dropping Z^*. The basic results were unaffected.

Salant, Walter S., "The Dollar and World Liquidity—A Minority View," *The Economist*, February 5, 1966, *218*, 526–29.

Fair, Ray C., "The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First Order Serially Correlated Errors," *Econometrica*, May 1970, *38*, 507–516.

Hood, Neil and Young, Stephen, *The Economics of Multinational Enterprise*, London: Longman, 1979.

Hufbauer, Gary C., "The Multinational Corporation and Direct Investment," in Peter B. Kenen, eds., *International Trade and Finance*, Cambridge: Cambridge University Press, 1975, 253–319.

Kenen, Peter B., *A Model of the U.S. Balance of Payments*, Lexington: Lexington Books, 1978.

Kindleberger, Charles P., *American Business Abroad: Six Lectures on Direct Investment*, New Haven: Yale University Press, 1969.

Kouri, Pentti J. K. and Porter, Michael G., "International Capital Flows and Portfolio Equilibrium," *Journal of Political Economy*, May/June 1974, *82*, 443–67.

Krueger, Russell C., "U.S. International Transactions, First Quarter 1981," *Survey of Current Business*, June 1981, 31–36.

Nelson, Charles R., *Applied Time Series Analysis*, San Francisco: Holden-Day, 1973.

Obstfeld, Maurice, "Exchange Rates, Inflation, and the Sterilization Problem," *European Economic Review*, March/April 1983, *21*, 161–89.

Ruffin, Roy J., "International Factor Movements," in Ronald W. Jones and Peter B. Kenen, eds., *Handbook of International Economics*, Vol. I, Amsterdam: North-Holland, 1984.

Tobin, James, "A General Equilibrium Approach to Monetary Theory," *Journal of Money, Credit, and Banking*, February 1969, 115–29.

The Nonequivalence of Tariffs and Import Quotas

By JAMES R. MELVIN*

It is now well known that in a world which is not perfectly competitive, tariffs and quotas are not equivalent. Jagdish Bhagwati (1965) showed that in the presence of monopoly the effects of tariffs and quotas could be quite different, and this argument was extended by Hirofumi Shibata (1968) and Bhagwati (1968). It is also well known that in the presence of uncertainty, or if there are changes in the economy such as growth or price fluctuations, tariffs and quotas may produce different results (see Gideon Fishelson and Frank Flatters, 1975, and Leslie Young, 1979). Carlos Rodriguez (1974) has also shown that tariffs and quotas produce different final equilibria if there is retaliation, and Edward Tower (1975) has analyzed the effects of imposing the optimum quota in combination with a variety of other trade restrictions. Tower finds that in the standard retaliation game, the volume of trade will approach zero, and he concludes that quantitative restrictions will never be preferred to tariffs in such situations. In particular, retaliatory behavior with quotas will always lead both countries towards the zero trade position, whereas retaliation with tariffs may lead to an equilibrium in which trade exists. Rodney Falvey (1975) has shown that, even in a perfectly competitive world, if the foreign offer curve is inelastic, some tariffs cannot be duplicated by an import quota.

What does not seem to be generally recognized is that, even in the static, perfectly competitive world, tariffs and import quotas are *never* equivalent if both countries pursue restrictive trade practices. It can also be shown that the form of trade restrictions used by one country can have significant implications for the range of instruments available to its trading partner.

The analysis uses the traditional two-country general-equilibrium model of international trade. The standard offer curve equilibrium is shown in Figure 1 where O_h and O_f are the offer curves for countries H and F, respectively. Both countries are assumed to produce two goods with production functions which exhibit constant returns to scale. Consumer preferences are represented by a set of community indifference curves, and the offer curves are shown to be inelastic throughout their range, although these assumptions are not required for the analysis. The free-trade equilibrium is at A with prices P. If both countries impose tariffs, the offer curves could shift to O_{ht} and O_{ft} with a new equilibrium at B.[1]

The question now is whether this tariff equilibrium, B, could be generated by import quotas. In Figure 2, the same offer curves and equilibrium price lines are shown, and now OJ and OK are the import quotas imposed by countries H and F, respectively. The offer curves for the two countries have now become $ONEJ$ for country H and $OMGK$ for country F. While B, the interior intersection of these two offer curves, is in equilibrium, it is unstable and therefore should not be sustained. This unstable equilibrium is bound by the two stable equilibria N and M which are associated with equilibrium prices P_1 and P_2, respectively. Which of these two equilibria will prevail depends on which country imposes the quota first. Note that only one quota can be effective and the effective quota will be the first one imposed.

---

*Department of Economics, Social Science Centre, University of Western Ontario, London, Canada N6A5C2. I am indebted to Bill Ethier, Ian Wooton, and a referee for helpful comments.

[1] Point B is shown on the same price line P simply to avoid clutter in the diagram. Any point in the interior of the lens-shaped area formed by the intersection of the two offer curves is possible.

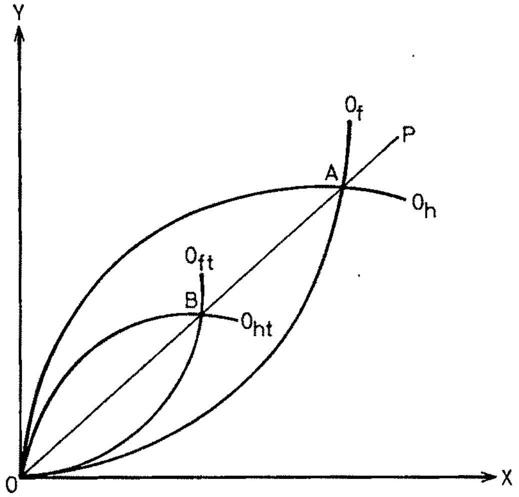


FIGURE 1

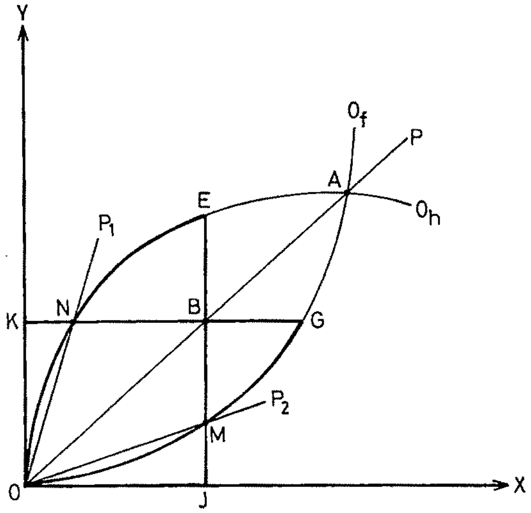


FIGURE 2

The adjustment mechanism assumed here is the usual one. It is assumed that the governments auction the quota rights and that the winning intermediators determine the commodity price which will clear the domestic markets. The world terms of trade is determined by the Walrasian auctioneer who calls out prices to which the intermediators respond with the quantities found from the quota-distorted offer curves of Figure 2.

It is clear that the argument just presented applies equally well for any pair of tariffs imposed by the two countries. None of the equilibria associated with tariffs by both countries can be reproduced by import quotas. It is interesting to note that while any point in the lens-shaped area *OEAG* can be generated by an appropriate choice of tariffs, only the boundary of this set can be generated by import quotas. Of course, with tariffs in both countries, none of the boundary points are possible in a tariff regime. Thus the feasible sets of equilibria for the tariff and quota cases are disjoint.

The fact that no tariff equilibria can be duplicated by import quotas depends on the fact that tariffs have been assumed to exist in both countries. If only one country imposes a tariff, then all equilibrium points will be located along the offer curve of the other country, or in other words, will be on the boundary of the set *OEAG*. It is therefore only in the special case where only one country is assumed to have a tariff that import quotas and tariffs are equivalent.

In numerical general-equilibrium analysis, it is often difficult to quantify the effects of quotas and other nontariff barriers to trade. One approach that is sometimes used is to calculate the tariff equivalents of such trade restrictions and to use these in the calculations. The analysis here suggests that such an approach may result in serious misspecifications of the model.

In the preceding, I have been careful to specify that only import quotas are being considered. With export quotas, or as they are now more commonly called, voluntary export restraints, the situation is quite different. Again consider Figure 2, but now assume that country *F* imposes a voluntary export restraint equal to *OJ* and that country *H* imposes a voluntary export restraint equal to *OK*. The offer curves for countries *H* and *F* become *ONG* and *OME*, respectively, and this gives rise to equilibrium at point *B*. This equilibrium is stable and thus any tariff regime can be duplicated by a set of voluntary export restraints. Indeed, voluntary export restraints can generate all the points both in the interior and on the boundary of the area *OEAG*.[2]

It is generally argued that voluntary export restraints are an inferior form of protection when compared either to tariffs or export quotas, because voluntary export constraints involve a transfer of revenue to foreign countries which would otherwise be captured by the domestic economy. Again this argument is seen to be true only in the case where voluntary export restraints are in place in one country but not the other. The voluntary export restraints of Figure 2 produce exactly the same equilibrium as the import tariffs in Figure 1, although domestic prices will generally not be the same in these two cases. Furthermore, voluntary export restraints can be shown to be superior to import quotas in a variety of circumstances. Voluntary export restraints permit a much larger number of possible equilibrium points, and this may be important if countries wish to cooperate in setting their commercial policies. Suppose that tariffs cannot be used, perhaps for political reasons, and that the choice is between quotas and voluntary export restraints. With the quotas of Figure 2, the only possible equilibrium process are P_1 or P_2, and, in either case, one of the two countries will be seriously disadvantaged by this form of trade restriction. With voluntary export restraints, trade can be restricted but the initial terms of trade, *P*, can be retained so that neither country is subjected to a serious terms-of-trade deterioration. Perhaps the current use of voluntary export restraints is not as misguided a policy action as most economists have assumed.

It is now clear why Rodrigues and Tower found that retaliation through quotas produced different results than retaliation with tariffs. Optimal quota policy always involves a point on the boundary of the lens-shaped area *OEAG* of Figure 2, since the highest attainable trade indifference curve will always go through a point such as *E*. Thus retaliation will lead inexorably towards the origin. Optimal policy with tariffs will always imply an equilibrium in the interior of *OEAG* (as long as both countries have tariffs) and an equilibrium with positive trade is possible. Indeed, point *B* of Figure 1 could be an equilibrium with retaliation.

Figure 2 also illustrates the fact that once one country has opted for quotas, the policy choices available to the other country are limited. Suppose that the foreign country has imposed an import quota resulting in the offer curve *OMGK*. All possible equilibria are restricted to the line segment *NG*, and only through the use of a tariff by country *H* can any point on the interior of *NG* be obtained. This is a special case of the Falvey result, where it is shown that with an inelastic offer curve some tariff equilibria cannot be duplicated by import quotas. An initial quota by country *F* results in an inelastic offer curve over the range *GK*. The Tower result that import quotas never dominate tariffs when retaliation is introduced is also illustrated by this example.

The conclusions I have derived depend on the instability of interior solutions when import quotas are imposed, and one might wonder why this result has not been noted in the earlier literature. One possible explanation is that much of the analysis comparing tariffs and quotas has been done using partial-equilibrium analysis, where these instability problems do not reveal themselves. We have, then, another reminder that general-equilibrium questions should be addressed using general-equilibrium techniques.

REFERENCES

Bhagwati, Jagdish N., "On the Equivalence of Tariffs and Quotas," in R. E. Baldwin et al., eds., *Trade, Growth and the Balance of Payments*: *Essays in Honor of Gottfried Haberler*, Chicago: Rand McNally, 1965.

______, "More on the Equivalence of Tariffs and Quotas," *American Economic Review*, March 1968, *58*, 142–46.

Falvey, Rodney E., "A Note on the Distinction

---

[2] One might be tempted to argue that voluntary export restraints (*VER*s) are the more general form of trade restriction, since they can reproduce any situation generated either by tariffs or import quotas. While technically correct, this argument is true only because some *VER*s are not binding. Boundary points can only be obtained if only one *VER* is effective. The point is that while all tariffs are effective, all *VER*s are not.

Between Tariffs and Quotas," *Economica*, August 1975, *42*, 319–26.

Fishelson, Gideon, and Flatters, Frank, "The (Non)Equivalence of Optimal Tariffs and Quotas Under Uncertainty," *Journal of International Economics*, November 1975, *5*, 385–93.

Rodriguez, Carlos Alfredo, "The Non-Equivalence of Tariffs and Quotas Under Retaliation," *Journal of International Economics*, August 1974, *4*, 295–98.

Shibata, Hirofumi, "A Note on the Equivalence of Tariffs and Quotas," *American Economic Review*, March 1968, *58*, 137–42.

Tower, Edward, "The Optimum Quota and Retaliation," *Review of Economic Studies*, October 1975, *42*, 623–30.

Young, Leslie, "Ranking Optimal Tariffs and Quotas for a Large Country Under Uncertainty," *Journal of International Economics*, May 1979, *9*, 249–64.

The Welfare Economics of Price Supports in U.S. Agriculture

By ERIK LICHTENBERG AND DAVID ZILBERMAN*

Government regulations aimed at furthering social goals such as environmental quality and equitable income distribution often reduce productivity. Estimating these costs of regulation is one of economists' main contributions to the policy evaluation process. Assessments of the economic impacts on product markets are generally based on an assumption of competitive market clearing. As Arnold Harberger (1971) pointed out, this assumption is inappropriate when market distortions are present. Preexisting government programs are an important source of such distortions. Because they may have profound effects on resource allocation—and hence on the impacts of new regulations—in the markets affected, their presence makes necessary modification of the standard welfare analytic procedures.

This paper examines the welfare effects of regulation on product markets affected by revenue-support programs (price supports, marketing orders, and import quotas). Because monetary estimates of regulatory benefits are often not used in policy assessments and because in any event benefits tend to be evaluated separately from costs, we focus only on the latter. In addition, we deal only with economic losses realized in product markets and ignore components of total regulatory cost such as monitoring and enforcement costs. Section I presents a conceptual framework, and Section II applies it to the case of regulation restricting productivity of three major U.S. crops falling under current support programs.

I. Welfare Evaluation in Revenue-Supported Industries

Mechanisms such as price supports, import quotas, and marketing orders are used in a number of important industries—notably, those with low elasticities of demand, like agriculture—to support producers' incomes. These policies create situations where the effective price received by producers exceeds consumers' willingness to pay as measured by the demand curve. This gap may arise either from explicit government subsidies, or from government policies that effectively increase market-clearing prices above competitive levels. In the case of agricultural price supports, the difference between average revenue and the demand price represents an explicit subsidy paid out of tax revenues. Agricultural marketing orders essentially set up discriminating monopoly schemes for allocating an industry's output among markets so that the pooled price (average revenue) producers receive exceeds the demand price. Import quotas effectively increase domestic excess demand above competitive levels, that is, above the levels specified by the difference between domestic demand and foreign supply.

To simplify the exposition, we conduct the analysis in terms of one of the most important revenue-support programs—the target price mechanism which has been the mainstay of U.S. farm programs in recent years. The results obtained for this case can be generalized to other forms of revenue support as well (see our earlier paper, 1985). The target price scheme operates essentially as follows. Output is sold at the market price; when the market price falls below the target price, producers receive from the government a subsidy (known as the deficiency payment) equal to the difference between the two. When the competitive market-clearing price is below the target price, the target price becomes the effective price received by producers and thus the relevant price for production decisions, while output will be absorbed by consumers at a market price determined by the demand curve.

---

*Director of Environmental Economics, Western Consortium for the Health Professions, San Francisco, CA, and Associate Professor, Department of Agricultural and Resource Economics, University of California, Berkeley, CA. We thank Al McGartland and an anonymous referee for their helpful suggestions. This research was supported by the U.S. Environmental Protection Agency under assistance agreement CR811200-02. The views expressed are our own and not those of the Agency. (For reprint identification only, Giannini Foundation paper no. 796.)

Suppose that an agency imposes a new, productivity-decreasing regulation on a market with this form of price supports. The agency will typically have little or no influence over policies outside of its own specific jurisdiction regardless of whether such policies impinge on matters which concern it. The U.S. Environmental Protection Agency (EPA), for example, has jurisdiction over pesticide use, but not over preexisting policies such as the agricultural policy that exerts influence on pesticide-use patterns. Thus, the agency is forced to behave as a *policy taker* and, in evaluating the economic costs of regulation, must consider any prior policies as given.

The market welfare effects of this new regulation can be analyzed using the framework developed by T. D. Wallace (1962), Bruce Gardner (1983), and Richard Just (1986). As Figure 1 illustrates, the new regulation causes a leftward shift in the supply curve from S_1 to S_2, reducing output from Q_1 to Q_2 and raising the equilibrium price to the consumer from P_1 to P_2. By reducing output and increasing price, regulation causes a loss in consumers' surplus equal to the area $g+h+i+j+k+l$. Because the target price is unaffected by the regulation, the loss in producers' surplus equals the area $a+b+c$. Expenditures on the deficiency payment program decrease by an amount equal to the area $c+e+f+g+h+i+j+k+l$. One part of the savings in government expenditure, $g+h+i+j+k+l$, represents reductions in subsidies to consumers, while another part, c, represents reductions in subsidies to producers; neither has any net effect on social welfare. Thus, the net change in social welfare is $e+f-(a+b)$. Following Harberger, one can also interpret this market welfare effect as the sum of two other components: The net social loss the regulation would impose under competitive equilibrium, $-a$, plus the reduction in deadweight loss from $d+e+f$ to $b+d$, a net change of $e+f-b$. Thus, $e+f-a-b$ represents the net regulation-induced change on the social efficiency of production.

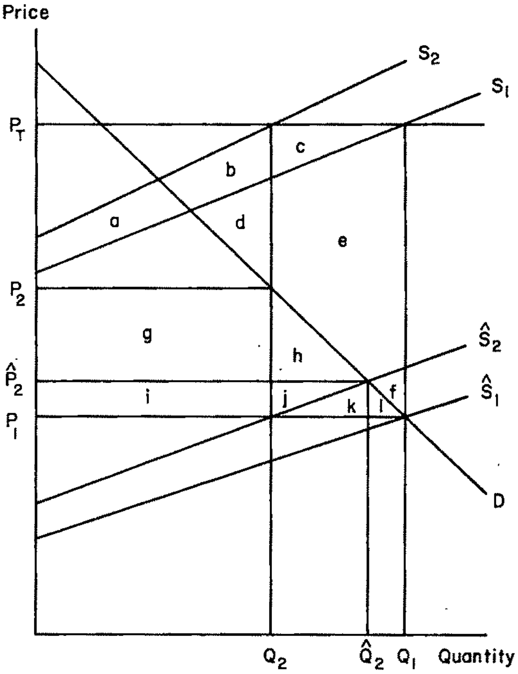


FIGURE 1. WELFARE IMPACTS OF REGULATION IN AN INDUSTRY WITH PRICE SUPPORTS

Note that, under price supports, the net effect of regulation on consumers in their dual role of consumers/taxpayers is always positive while the effect on producers is always negative. Industries with price supports generally have inelastic demand curves (low elasticities of demand have long been cited as a major motivation for agricultural price supports) which implies that, in the absence of price supports, the net effect of regulation on consumers would be negative and the net effect on producers would be positive. Thus, the distribution of gains and losses from regulation under price supports will tend to be the exact opposite of the situation under competitive equilibrium. For this reason, price supports can be said to strengthen producers' opposition to and consumers' support for regulation.

Welfare analyses are typically performed under the assumption that markets are in competitive equilibrium before and after regulatory measures are taken. This approach was employed by the National Academy of Sciences (1980) in its assessment of economic losses from bans on uses of the pesticide chlordimeform and in recent studies undertaken at the U.S. Department of Agriculture (see, for example, Glenn Collins and C. Robert Taylor, 1983, and Craig Osteen and Fred Kuchler, 1986, among others). When the industry in question has effective price supports, following this procedure means using a misspecified model; and one can expect the standard methods to produce biased welfare estimates. The preceding discussion suggests that this specification error should have no effect on estimation of the consumer demand curve. Thus, one would expect an analysis performed under this misspecification to use the true demand curve $D(Q)$. Estimation of the supply curve under the assumption of competitive equilibrium and calibration to preregulation market price and quantity, though, should produce a curve like $\hat{S}_1(Q)$ in Figure 1 which lies everywhere below the true supply curve and which intersects the demand curve at (Q_1, P_1). If the regulation was assumed to have the same effect on the misspecified supply curve as on the true supply curve $(\hat{S}_2 - \hat{S}_1 = S_2 - S_1)$, then the traditional approach would calculate the market welfare effects based on a shift in the competitive equilibrium quantity and price from (Q_1, P_1) to $(\hat{Q}_2, \hat{P}_2)$. It is evident from Figure 1 that, because the traditional approach underestimates the impact of regulation on output and price, it always underestimates consumers' losses from regulation (in Figure 1 this bias equals the area $g+h$). It is not hard to show that, for the same reason, the traditional approach usually underestimates producers' losses as well (in Figure 1, S_1 and $\hat{S}_1$ are parallel so that the bias equals the area $i+j$). However, because it ignores the positive impact of reductions in government revenue-support program payments, the traditional approach always overestimates net market losses from regulation (in Figure 1, this bias equals the area $c+e+f+k+l$). Thus, the traditional approach can be said to be biased against regulation.

II. An Empirical Example

These effects were illustrated using the case of a regulation affecting production of three important U.S. crops covered by agricultural support programs: corn, cotton, and rice. Inverse supply and demand for all three were assumed to have constant elasticity forms $S_1(Q) = b_0 Q^{b_1}$ and $D(Q) = a_0 Q^{-a_1}$. The parameters used to construct these curves are shown in Table 1. Elasticities of supply and demand were chosen from studies reflecting consensus values. The parameters a_0 and b_0 were calibrated to equate inverse demand with the market price and inverse supply with the target price at current output levels using estimates of output, target prices, and market prices chosen to be broadly representative of the levels prevailing in recent years. The misspecified inverse supply curve $\hat{S}_1(Q)$ reflecting the assumption of competitive equilibrium prior to the pesticide ban was assumed to have the same elasticity as the true curve but a different constant term, $\hat{b}_0$, which was chosen to equate inverse supply with the market price at current output levels. The regulations were assumed to decrease the supply of these crops by 1 percent—a figure which appears to be not atypical of pesticide bans (for example, National Academy of Sciences, 1983)—so that $S_2(Q) = 1.01\ b_0 Q^{b_1}$ and $\hat{S}_2(Q) = 1.01\ \hat{b}_0 Q^{b_1}$. The model was then used to calculate the welfare effects of the ban with and without price supports and the welfare effects estimated under an incorrect assumption of competitive equilibrium when price supports were present.

The results of this procedure, shown in Table 2, demonstrate that 1) the savings in U.S. Treasury costs from regulation are quite substantial, 2) price supports cause sizable changes in the welfare effects of regulation, 3) regulation will have no net market effect for target prices not too much greater than those currently in force, and 4) treating the behavior observed under price supports as if

TABLE 1—MODEL PARAMETERS

Crop	Elasticity of Supply	Elasticity of Demand	Output	Market Price	Target Price
Corn[a]	.30	−.50	8.00	2.50	3.00
Cotton[b]	.50	−.41	12.00	.60	.80
Rice[c]	.80	−.43	.15	9.00	11.90

Sources: For corn supply and demand see, Bruce Gardner (1984); for cotton supply, see Terry Townsend, Carol Skelly, and Phil Sronce (1984); for cotton demand, see Townsend (1985); for rice supply, see Keith Collins and Sam Evans (1982); for rice demand, see Warren R. Grant, John Beach, and William Lin (1984); and for output, market prices, and target prices, see U.S. Department of Agriculture (1983).

[a] Output measured in billion bushels and price in dollars per bushel.

[b] Output measured in million bales and price in dollars per pound.

[c] Output measured in billion hundredweight and price in dollars per hundredweight.

TABLE 2—WELFARE EFFECTS OF A PESTICIDE BAN[a]

	ΔCS	$\Delta \hat{C}S$	ΔPS	$\Delta \hat{P}S$	ΔGP	$\Delta CS - \Delta GP$	ΔW	$\Delta \hat{W}$
Corn								
Competitive Equilibrium	−77.30		29.73				−47.57	
Target Price (dollars per bushel)								
3.00[b]	−119.58	−74.70	−55.03	28.73	−131.33	11.74	−43.28	−45.97
3.50	−114.18	−71.32	−67.24	29.13	−144.50	30.32	−36.92	−46.19
4.00	−109.69	−68.52	−79.98	21.73	−158.82	49.13	−30.85	−47.79
4.50	−105.89	−66.14	−93.22	18.47	−174.13	68.24	−24.98	−47.67
5.00	−102.59	−64.09	−106.90	15.30	−190.26	87.67	−19.23	−48.78
5.50	−99.70	−62.28	−121.00	12.19	−207.15	107.45	−13.55	−50.09
6.00	−97.13	−60.67	−135.49	9.11	−224.71	127.58	−7.92	−51.56
6.50	−94.83	−59.23	−150.35	6.07	−242.87	148.04	−2.31	−53.17
Cotton								
Competitive Equilibrium	−43.28		17.02				−26.26	
Target Price (dollars per pound)								
0.80[b]	−87.68	−39.43	−31.76	15.51	−99.38	11.69	−20.07	−23.92
0.90	−80.56	−36.22	−37.90	12.12	−104.38	23.82	−14.08	−24.10
1.00	−74.68	−33.58	−44.39	8.97	−110.64	35.97	−8.42	−24.60
1.10	−69.73	−31.35	−51.21	6.00	−117.96	48.23	−2.98	−25.36
1.20	−65.50	−29.45	−58.35	3.13	−126.17	60.67	2.32	−26.32
1.30	−61.83	−27.80	−65.79	0.34	−135.17	73.34	7.55	−27.46
1.40	−58.62	−26.36	−73.53	−2.40	−144.88	86.26	12.73	−28.76
1.50	−55.78	−25.08	−81.55	−5.11	−155.24	99.45	17.91	−30.19
Rice								
Competitive Equilibrium	−9.71		3.07				−6.63	
Target Price (dollars per hundredweight)								
11.90[b]	−25.12	−8.75	−7.86	2.77	−28.47	3.35	−4.51	−5.98
12.30	−24.26	−8.45	−8.34	2.46	−28.85	4.59	−3.76	−5.99
12.70	−23.45	−8.17	−8.84	2.16	−29.28	5.83	−3.01	−6.01
23.10	−22.69	−7.91	−9.34	1.87	−29.76	7.07	−2.28	−6.04
13.50	−21.98	−7.66	−9.87	1.58	−30.29	8.31	−1.56	−6.08
13.90	−21.31	−7.42	−10.40	1.30	−30.85	9.56	−0.84	−6.12
14.30	−20.68	−7.20	−10.94	1.02	−31.49	10.81	−0.14	−6.18
14.70	−20.08	−7.00	−11.50	0.74	−32.15	12.07	0.57	−6.25

Note: Shown in million dollars.

[a] ΔCS, ΔPS, ΔGP, and ΔW denote changes in consumers' surplus, producers' surplus, government payments, and net welfare, respectively; $\Delta \hat{C}S$, $\Delta \hat{P}S$, and $\Delta \hat{W}$ denote changes in consumers' surplus, producers' surplus, and net welfare, respectively, calculated under the erroneous assumption of competitive equilibrium.

[b] Denotes current target price.

it were generated by competitive equilibrium leads to substantial overesimation of the net market effect of regulation.

First, note that the savings in U.S. Treasury costs associated with regulation are quite high for all of these crops. At current target price levels, for example, the ratio of government savings to the sum of producers' and consumers' losses ranges from about .75 [131.33/(119.58 + 55.03)] for corn to roughly .85 for cotton and rice. Consequently, it is apparent that ignoring this component of market welfare will lead to serious distortions in evaluating policy alternatives.

Second, compare the welfare effects of regulation in competitive equilibrium with those with target prices at current levels. In competitive equilibrium, producers gain substantially from regulation—$30 million for corn producers, $17 million for cotton growers, and $3 million for rice growers. As a result, one would expect members of this group (or at least some sections of it) to support regulation. Consumers, on the other hand, sustain substantial losses from the ban in every case—$10 million for rice, $43 million for cotton, and $77 million for corn. (In the case of pesticides, for example, one would thus expect only the more environmentally minded members of this group to support regulation.) With price supports, the situation is reversed: Consumers in their dual role as consumers/taxpayers benefit unambiguously (and considerably) to the tune of $3 million for rice and $12 million for corn and cotton and can thus be expected to support regulation more wholeheartedly, while producers suffer rather large losses—about $8 million for rice, $32 million for cotton, and $55 million for corn. This suggests that agricultural policy is a major cause of farmers' opposition to environmental regulation; without it, farmers might well support environmental regulation in many cases.

The third point of interest is that, for some crops, regulation will have no net market effect at target prices not much greater than those currently in effect: For rice, the zero-impact target price is only $2.48 per hundred weight (about 20 percent) higher than the current target price. One can see that the net market effect of regulation is both lower and declines more rapidly as supply becomes more elastic. For example, corn, cotton, and rice have roughly the same elasticities of demand (.4 to .5), while rice has a supply elasticity nearly twice that of cotton and about three times that of corn. Correspondingly, the proportional increase in the target price of rice required to produce a zero net market impact from the ban is only about one-half that required for cotton and only about one-sixth that required for corn. These results imply that environmental regulation becomes more attractive as government subsidization of agriculture grows, especially in markets where supply is less inelastic.

Finally, it is evident that treating markets with price supports as if they are in competitive equilibrium leads to significant overestimation of the net economic costs of regulation. There are two ways of measuring welfare under the erroneous assumption of competitive equilibrium. The first uses the true inverse supply curve (the "textbook model"); the second uses the inverse supply curve which gives market clearing at actual prices and quantities (the "estimated model"). In either case, the erroneous estimates of net market losses from regulation are substantially higher than the true losses at current target price levels. Under the textbook model, the overestimate is about 10 percent for corn (47.57 vs. 43.28), over 30 percent higher for cotton, and almost 50 percent higher for rice. Under the estimated model, the overestimate is about 6 percent for corn (45.97 vs. 43.28), close to 20 percent higher for cotton, and about 33 percent higher for rice. The size of the bias clearly increases as the elasticity of supply increases: it is smallest for the most inelastic crop (corn), larger for the less inelastic (cotton), and largest for the least inelastic (rice). Thus, the use of a competitive equilibrium framework may introduce serious errors into regulatory welfare analyses—errors which may well be serious enough to alter policy recommendations. Producers whose supply is less inelastic are of particular concern in this regard.

These results have some interesting implications for environmental regulation in agriculture. Cotton and rice are implicated in many environmental problems because of their intensive use of water and pesticides. As Table 2 indicates, for these crops, the distortional impacts of government programs are relatively large. For this reason, the net market impact of new regulation tends to be small (relative to the impacts on producers and consumers). Furthermore, the bias in estimating net market welfare under the erroneous assumption of competitive equilibrium tends to be substantial.

III. Conclusion

Government regulation has become sufficiently pervasive that new regulations are impinging on preexisting policies to an increasing extent. Evaluations of the market effects of newer regulations must, therefore, take account of these prior regulations. This paper has developed a framework for analyzing the market welfare effects of regulation for the case of a preexisting regime of revenue-support programs. To facilitate the analysis, several simplifications were made. We ignored the specific secondary features many revenue-support programs have (for example, the set-aside requirements of agricultural price-support programs), any feedback effects regulation may have on prior policies (for example, environmental regulation in agriculture may increase production costs and thereby induce increases in target prices since the latter are pegged to costs), etc. Nevertheless, the presence of these revenue-support programs was shown to cause striking alterations in the welfare effects of new regulations; moreover, these changes were significant enough both qualitatively and quantitatively to indicate that ignoring prior policies introduces serious distortions into regulatory welfare analyses. Price supports were shown to have large effects on the distribution of costs in the majority of cases turning significant gains for producers into sizable losses and sizable losses to consumers into noticeable gains. For some crops, the target price at which regulation had a zero market welfare impact turned out to be not too much higher than the levels currently in effect. The erroneous use of a competitive equilibrium framework was shown to introduce significant biases into estimates of the market welfare effects of regulation: For some crops, the upward bias was on the order of one-third to one-half of the actual market welfare effect. It is thus evident that ignoring revenue-support programs in performing welfare analyses may produce seriously distorted policy recommendations.

REFERENCES

Collins, Glenn S. and Taylor, C. Robert, "TECHSIM—A Regional Field Crop and Livestock Econometric Simulation Model," *Agricultural Economics Research*, April 1983, *35*, 1–8.

Collins, Keith and Evans, Sam, "Participation Incentives and Supply Effects of the 1982 Rice Average Reduction Program," *Rice Outlook and Situation*, March 1982, 9–12.

Gardner, Bruce, "Policy Options for Grains," in Gordon C. Rausser and Kenneth R. Farrell, eds., *Alternative Agricultural and Food Policies and the 1985 Farm Bill*, Berkeley: Giannini Foundation of Agricultural Economics, 1984, 81–99.

______, "Efficient Redistribution through Commodity Markets," *American Journal of Agricultural Economics*, May 1983, *65*, 225–34.

Grant, Warren R., Beach, John and Lin, William, "Factors Affecting Supply, Demand and Prices of U.S. Rice," ERS Staff Report No. AGES84083, Washington: U.S. Department of Agriculture, October 1984.

Harberger, Arnold C., "Three Basic Postulates for Applied Welfare Economics: An Interpretive Essay," *Journal of Economic Literature*, September 1971, *9*, 785–97.

Just, Richard E., "Automatic Adjustment Rules for Agricultural Policy Controls," American Enterprise Institute, Washington, 1986.

Lichtenberg, Erik and Zilberman, David, "The Welfare Economics of Regulation in Revenue-Supported Industries," Department of Agricultural and Resource Economics, University of California-Berkeley, 1985.

Osteen, Craig and Kuchler, Fred, "Potential Bans of Corn and Soybean Pesticides," Agricultural Economic Report No. 546, U.S. Department of Agriculture, Washington, April 1986.

Townsend, Terry, *Internal Estimates of Cotton Demand Elasticities*, Washington: U.S. Department of Agriculture, February 1985.

______, **Skelly, Carol and Sronce, Phil,** "Outlook for Participation in the 1985 Upland Cotton Program," *Cotton and Wool Situation and Outlook*, December 1984, 15–20.

Wallace, T. D., "Measures of Social Costs of Agricultural Programs," *Journal of Farm Economics*, December 1962, *44*, 580–94.

National Academy of Sciences, *Regulating Pesticides*, Washington, 1980.

______, *Pest Control for Corn and Soybeans*, Washington, 1983.

U.S. Department of Agriculture, *Agricultural Statistics*, Washington, 1983.

The Persistence of Volatility and Stock Market Fluctuations

By JAMES M. POTERBA AND LAWRENCE H. SUMMERS*

Explaining the dramatic intertemporal variation in real stock market prices is a fundamental problem of financial economics. Between 1960 and 1985, a relatively tranquil period by historical standards, the Dow Jones Industrial Average measured in 1985 dollars varied from 3385 (June 1966) to 909 (June 1982).[1] As the analysis of William Brainard, John Shoven, and Laurence Weiss (1980) suggests, it is difficult to account for the movements in the stock market during the last twenty years on the basis of changes in either expected cash flows or real interest rates. Robert Shiller (1981) provides significant evidence that stock market volatility in general cannot be explained by movements in the rational expectation of future dividends and interest rates.[2] A natural alternative hypothesis is that market movements reflect in substantial part changing risk premia induced by movements in stock market volatility. This explanation has been suggested by Burton Malkiel (1979) and Robert Pindyck (1984) in connection with the significant decline in real stock market values between the mid-1960's and the early 1980's. They argue that changes in risk are responsible for a significant part of the variation in share prices.

This paper evaluates the changing risk premium hypothesis and examines the influence of changing stock market volatility on the level of stock prices. We find that shocks to volatility decay rapidly and therefore can affect required returns for only short intervals. This implies that volatility shocks can have only a small impact on stock market prices, and makes us skeptical of the claim that volatility-induced fluctuations in risk premia account for much of the observed variation in these prices.

The paper is divided as follows: Section I clarifies the theoretical relationship between return volatility, required rates of return, and the level of share prices. Section II examines the time-series properties of stock market volatility using daily return data on the Standard and Poor's Composite Stock Index for the period 1928–84. The results suggest that while volatility is serially correlated, changes in current volatility have relatively small effects on volatility forecasts over even short horizons. Section III uses a different type of data, the implied volatilities in option premia, to examine the persistence of volatility shocks. The results again suggest that these shocks are only weakly serially correlated. The concluding section discusses the implications of our results for alternative explanations of recent stock market movements.

I. Volatility, Required Returns, and Stock Price Fluctuations

This section explores the relationship between changes in stock market return volatility and changes in the level of stock market prices. We assume that firms are not levered and that expected dividends grow at a constant rate. The former assumption allows us to ignore Fischer Black's (1976) important observation that the level of share prices, by affecting the degree of leverage, should have a direct impact on volatility. The latter assumption is maintained for convenience and could be relaxed easily. To parallel our empirical work, we use a discrete time formulation.

---

*Departments of Economics, MIT, Cambridge, MA 02139 and Harvard University, Cambridge, MA 02138, respectively. We are indebted to David Cutler, Ignacio Mas, and Carla Ponn for research assistance, and to Fischer Black, Olivier Blanchard, Zvi Griliches, Greg Mankiw, Robert Merton, Daniel Nelson, Julio Rotemberg, and two referees for helpful suggestions. Financial support was provided by the National Bureau of Economic Research and the National Science Foundation. This research is part of the NBER's research program in Financial Markets and Monetary Economics.

[1] This calculation employs monthly averages of the daily closing values of the Dow Jones Industrial Average, adjusted to December 1985 prices using the Urban Worker Consumer Price Index.

[2] Shiller's claim that stock prices are too volatile has sparked substantial controversy. Allan Kleidon (1985) and Terry Marsh and Robert Merton (1986) dispute his claim, while N. Gregory Mankiw, David Romer, and Matthew Shapiro (1985) provide supporting evidence.

We assume that share prices satisfy the standard requirement that

$$\text{(1)} \qquad \frac{E_t(P_{t+1})-P_t}{P_t}+\frac{D_t}{P_t}=r_t+\alpha_t,$$

where r_t denotes the risk-free real interest rate, α_t the equity risk premium, and D_t the dividend paid in period t. Equivalently, equation (1) can be written

$$\text{(2)} \quad P_{t+1}=(1+r_t+\alpha_t)P_t-D_t+\varepsilon_t P_t,$$

where $\varepsilon_t=(P_{t+1}-E_t[P_{t+1}])/P_t$ is a random disturbance assumed to be uncorrelated with any information available at time t. It reflects revisions in expectations about future values of D, α, and r which take place between t and $t+1$.

Equation (2) can be solved forward subject to an appropriate transversality condition to yield an expression for P_t:

$$\text{(3)} \qquad P_t=E_t\left[\sum_{j=0}^{\infty} R_{t+j}D_{t+j}\right]$$

where $R_{t+j}=\prod_{i=0}^{j}(1+r_{t+i}+\alpha_{t+i})^{-1}$.

Assuming that the real risk-free rate is constant, this expression may be linearized around the mean value of the risk premium $\bar{\alpha}$ to obtain

$$\text{(4)} \quad P_t=\sum_{j=0}^{\infty}\frac{E_t[D_{t+j}]}{(1+r+\bar{\alpha})^{j+1}} + \sum_{j=0}^{\infty}\frac{\partial P_t}{\partial \alpha_{t+j}}\left(E_t[\alpha_{t+j}]-\bar{\alpha}\right),$$

where

$$\text{(5)} \quad \partial P_t/\partial \alpha_{t+j}=-(1+r+\bar{\alpha})^{-j-2} \times \sum_{k=0}^{\infty}\frac{E_t(D_{t+j+k})}{(1+r+\bar{\alpha})^{k}}.$$

Equation (4) expresses current stock prices as a linear function of expected future risk premia. If expected dividends grow at a constant rate g, so that $E_t[D_{t+j}]=(1+g)^jD_t$, then (5) simplifies to

$$\text{(6)} \quad \frac{\partial P_t}{\partial \alpha_{t+j}}=\frac{-D_t(1+g)^j}{(1+r+\bar{\alpha})^{j+2}} \times \sum_{k=0}^{\infty}\frac{(1+g)^k}{(1+r+\bar{\alpha})^k} = \frac{-D_t(1+g)^j}{(1+r+\bar{\alpha})^{j+1}(r+\bar{\alpha}-g)}.$$

We now consider the link between return volatility and the equity risk premium. Robert Merton (1973, 1980) derives a linear relationship between the equity risk premium, α_t, and the variance of equity returns, σ_t^2:

$$\text{(7)} \qquad \alpha_t=\gamma\sigma_t^2.$$

The parameter γ is a function of investors' coefficients of relative risk aversion.[3] To study the effect of volatility changes on P_t, we must specify the evolution of σ_t^2. Our empirical work suggests that for the postwar period, monthly values of σ_t^2 follow an AR(1) process:

$$\text{(8)} \qquad \sigma_t^2=\rho_0+\rho_1\sigma_{t-1}^2+\mu_t.$$

It immediately follows that monthly values of α_t also follow an AR(1) process:

---

[3] In a one-consumer economy, γ equals the consumer's relative risk aversion. More generally, it is a weighted average of different consumers' relative risk aversions.

$$(9)\qquad \alpha_t = \gamma\rho_0 + \rho_1\alpha_{t-1} + \eta_t,$$

where $\eta_t = \gamma\mu_t$. The mean value of α_t is therefore $\bar{\alpha} = \gamma\rho_0/(1-\rho_1)$, and $(\alpha_t - \bar{\alpha})$ evolves according to

$$(10)\qquad \alpha_t - \bar{\alpha} = \rho_1(\alpha_{t-1} - \bar{\alpha}) + \eta_t.$$

From equation (10) we know that $E_t(\alpha_{t+j} - \bar{\alpha}) = \rho_1^j(\alpha_t - \bar{\alpha})$. This may be used along with equation (6) to simplify (4):

$$(11)\qquad P_t = \sum_{j=0}^{\infty} \frac{E_t(D_{t+j})}{(1+r+\bar{\alpha})^{j+1}} - \sum_{j=0}^{\infty} \frac{D_t(1+g)^j\rho_1^j(\alpha_t - \bar{\alpha})}{(1+r+\bar{\alpha})^{j+1}(r+\bar{\alpha}-g)}$$

$$= \frac{D_t}{r+\bar{\alpha}-g} - \left[\frac{1}{1+r+\bar{\alpha}-\rho_1(1+g)}\right] \times \left[\frac{D_t}{r+\bar{\alpha}-g}\right](\alpha_t - \bar{\alpha})$$

$$= \bar{P}_t + \frac{dP_t}{d\alpha_t}(\alpha_t - \bar{\alpha}).$$

The last expression shows the effect of risk premia shocks on share prices. Using the fact that $dP_t/d\sigma_t^2 = (dP_t/d\alpha_t)(d\alpha_t/d\sigma_t^2)$, we know

$$(12)\qquad \frac{dP_t}{d\sigma_t^2} = \frac{-\gamma}{[1+r+\bar{\alpha}-\rho_1(1+g)]} \times [D_t/(r+\bar{\alpha}-g)].$$

Multiplying both sides of (12) by σ_t^2/P_t yields

$$(13)\qquad \frac{d\log P_t}{d\log\sigma_t^2} = \frac{-\gamma\sigma_t^2\lambda_t(r+\bar{\alpha}-g)^{-1}}{[1+r+\bar{\alpha}-\rho_1(1+g)]}$$

where λ_t is the dividend yield, D_t/P_t. Since $\gamma\sigma_t^2 = \alpha_t$ and $\lambda_t = r_t + \alpha_t - g$, the derivative in (13) can be written as[4]

$$(14)\qquad \frac{d\log P_t}{d\log\sigma_t^2} = \frac{-\bar{\alpha}}{[1+r+\bar{\alpha}-\rho_1(1+g)]}.$$

The absolute value of this derivative rises with ρ_1. If increases in volatility are expected to persist, they will have a greater impact on the discount factors applied to future cash flows and therefore on current share prices.

To gauge the possible influence of volatility on share prices, we evaluate (14) at some plausible parameter values. Roger Ibbotson (1984) reports that the mean annual nominal return on common stocks for the period 1948–83 was 11.6 percent, or .964 percent per month. The mean nominal return on Treasury bills was 4.6 percent per year, implying $\bar{\alpha} = .006$ per month. The average real return on Treasury bills (r) was 0.4 percent per year or 0.035 percent per month. The estimated variance of market returns, expressed at monthly rates, ranged from .00022 in 1964 to .0053 in 1974 and averaged .0017. The last statistic, in conjunction with the estimate for the mean risk premium, implies $\gamma = 3.5$.[5] The annual growth rate of nominal dividends on the Standard and Poor's Composite Index during the 1948–83 period was 5.2 percent. Combining this with our annual inflation rate of 4.2 percent yields an average real annual dividend growth rate (g) of 0.01, or 0.00087 per month.

The effect of changes in volatility on the level of share prices is very sensitive to the level of ρ_1, the serial correlation in monthly volatility. The derivative in (14) equals $-.006$ when $\rho_1 = 0$, $-.012$ when $\rho_1 = .5$, $-.108$ when $\rho_1 = .9$, and $-.387$ when $\rho_1 = .99$. Stated another way, a doubling of volatility from its average level reduces the value of the market by only 0.6 percent if $\rho_1 = 0$, by 1.2 percent if $\rho_1 = .5$, and by 38.7 percent if $\rho_1 = .99$. These calculations show that if fluctuations in volatility are to play a significant role in explaining stock market fluctua-

---

[4] We obtain (14) by evaluating (13) at the mean return, dividend yield, and risk premium.

[5] Merton (1980) estimates γ to be 3.2 using data for the postwar period.

TABLE 1—AUTOCORRELOGRAM OF MONTHLY VOLATILITY SERIES

Lag Length	1950–84		1930–84	
	Autocorrelation	Partial Autocorrelation	Autocorrelation	Partial Autocorrelation
1	0.571 (0.040)	0.571 (0.040)	0.730 (0.027)	0.730 (0.027)
2	0.384 (0.045)	0.086 (0.049)	0.500 (0.031)	0.042 (0.035)
3	0.277 (0.047)	0.041 (0.049)	0.459 (0.029)	0.229 (0.034)
4	0.219 (0.048)	0.042 (0.049)	0.493 (0.029)	0.169 (0.033)
5	0.224 (0.048)	0.093 (0.049)	0.454 (0.030)	0.030 (0.034)
6	0.185 (0.048)	0.003 (0.049)	0.478 (0.029)	0.215 (0.032)
7	0.148 (0.048)	0.003 (0.049)	0.493 (0.029)	0.079 (0.033)
8	0.162 (0.048)	0.069 (0.049)	0.467 (0.029)	0.059 (0.033)
9	0.169 (0.048)	0.047 (0.049)	0.412 (0.030)	−0.008 (0.033)
10	0.174 (0.048)	0.038 (0.050)	0.414 (0.030)	0.069 (0.033)
11	0.132 (0.049)	−0.027 (0.050)	0.426 (0.030)	0.044 (0.032)
12	0.107 (0.049)	−0.003 (0.050)	0.419 (0.030)	0.029 (0.033)
13	0.109 (0.049)	0.027 (0.050)	0.369 (0.031)	−0.051 (0.033)
14	0.059 (0.048)	−0.049 (0.049)	0.326 (0.032)	−0.058 (0.033)
15	0.072 (0.048)	0.029 (0.048)	0.287 (0.032)	−0.067 (0.033)
16	0.065 (0.048)	−0.005 (0.049)	0.300 (0.032)	0.029 (0.033)
17	0.067 (0.048)	0.014 (0.049)	0.317 (0.032)	0.029 (0.033)
18	0.055 (0.048)	−0.015 (0.049)	0.282 (0.033)	−0.076 (0.033)
19	0.030 (0.048)	−0.025 (0.049)	0.246 (0.033)	−0.007 (0.033)
20	0.002 (0.048)	−0.034 (0.049)	0.251 (0.033)	0.020 (0.033)
21	−0.015 (0.048)	−0.024 (0.049)	0.242 (0.033)	−0.004 (0.033)
22	0.017 (0.048)	0.049 (0.049)	0.264 (0.033)	0.110 (0.032)
23	0.019 (0.048)	−0.003 (0.049)	0.295 (0.032)	0.071 (0.033)
24	0.010 (0.048)	−0.010 (0.049)	0.269 (0.033)	−0.021 (0.033)

Note: Monthly volatility estimates are calculated as the average value of squared daily returns on the Standard and Poor's Composite Stock Index. Standard errors are shown in parentheses.

tions, then ρ_1 must be quite large. The next two sections present estimates of this parameter.

II. Estimates of Serial Correlation in Market Volatility

This section examines the persistence of volatility over the 1928–84 period as well as the shorter postwar period studied by Pindyck. We use volatility estimates computed from daily returns on the Standard and Poor's (S&P) Composite Index, which were kindly provided to us by Kenneth French, G. William Schwert, and Robert Stambaugh.[6] The variance estimator for month t, $\hat{\sigma}_t^2$, is

$$\hat{\sigma}_t^2 = \sum_{i=1}^{K_t} \tilde{s}_{t,i}^2 / k_t, \tag{15}$$

where $\tilde{s}_{t,i}$ is the market return on the ith day of month t, measured by the percentage change in the S&P Composite Index, and k_t is the number of trading days in month t.

The autocorrelogram and partial autocorrelogram for the estimated volatility series are shown in Table 1. The first panel corresponds to the post-1950 period as in Pindyck, while the second utilizes data from the full sample period. The data clearly exhibit positive serial correlation, although the persistence of volatility is relatively unimportant from an economic perspective. For the postwar sample, the monthly first-order autocorrelation coefficient is .571. The partial autocorrelogram shows only one statistically significant value at lags higher than one month, that at lag 5, and suggests that an AR(1) representation provides an adequate description of the data.

---

[6] An earlier version of this paper used daily stock market returns for the 1968–84 period, as well as monthly returns for the 1871–1984 period, to investigate volatility persistence.

The volatility data for the full sample period display somewhat greater persistence. The first-order autocorrelation coefficient is 0.730, and the autocorrelogram decays much more slowly than that for the postwar period. The difference in persistence is primarily due to the presence of the Great Depression years in our 1930–84 sample. As Robert Officer (1973) observes, the volatility of market returns was much higher during the Great Depression than at any time in the last century. Moreover, since volatility was high for several consecutive months, these few observations have a significant impact on the estimated serial correlation parameters. It is far from clear that the persistence of volatility during the Great Depression provides a useful guide to the current persistence beliefs of market participants.

Our analysis of the autocorrelation properties of the level of volatility implicitly assumes that volatility is a stationary series. French, Schwert, and Stambaugh (1985), however, argue that volatility is actually nonstationary and fit ARMA models to monthly differences in volatility.[7] Following David Dickey and Wayne Fuller (1981) and Fuller (1976), we explicitly tested the hypothesis that measured volatility is a nonstationary series, that is, that there is a unit root in its autoregressive representation. For both sample periods, the null hypothesis of nonstationarity was rejected in favor of the alternative hypothesis of stationarity at very high confidence levels. For example, when we test the unit root hypothesis allowing for a relatively general AR(12) specification, the test values are -3.70 and -3.97 for the 1930–84 and 1950–84 samples, respectively, while the .01 critical values are -3.46 and -3.44. Similar results obtain for a wide range of different autoregressive lag lengths.[8]

Even if volatility is modeled as a nonstationary series, volatility shocks are not highly persistent. The results of estimating an IMA(1,3) process, as in French et al., for the full sample period are

$$\Delta\sigma_t^2 = (1 - \underset{(.038)}{.404}L - \underset{(.039)}{.327}L^2 - \underset{(.035)}{.048}L^3)\varepsilon_t.$$

Three months after a volatility shock, the level of volatility is elevated by only .221 times the initial shock. Although volatility remains at this new higher level forever, the rapid decay implied by the moving average parameters makes the net effect on share prices relatively small.

To illustrate what our results imply about the impact of volatility shocks on share prices, we used equation (4) to evaluate $d\log P_t/d\log\sigma_t^2$ for several different stochastic specifications of the volatility process. We use (*i*) an AR(1) process, which as noted above describes our data reasonably well; (*ii*) an AR(12) process designed to capture long-run persistence which might be poorly characterized by an AR(1) model; and (*iii*) the IMA(1,3) model described above. In each case we compute the impulse response functions for the level of volatility, and use (4) to calculate the change in share prices associated with each of these perturbations to the volatility path.

The resulting calculations, along with the standard errors of the estimated elasticities,[9] are shown in Table 2. The impact of a volatility shock on share prices is extremely

---

[7] French et al. model the square root of volatility, and in a revised version of their paper, the logarithm of volatility, as an IMA(1,3) process.

[8] Schwert (1985) suggests that Dickey-Fuller tests may yield spurious results if the time-series process is misspecified, although this problem is reduced when long autoregressive processes are considered. We therefore conducted unit root tests assuming up to AR(24) processes for volatility, and rejected the unit root hypothesis in all cases.

[9] We calculate the asymptotic variance of our elasticity estimates by computing numerical derivatives of these elasticities with respect to each of the estimated time-series parameters, stacking these derivatives in a column vector f, and then evaluating $f'\Omega f$ where Ω is the variance-covariance matrix for the parameter estimates.

TABLE 2—ESTIMATES OF ELASTICITY OF SHARE PRICES WITH RESPECT TO VOLATILITY

Estimation Period and Volatility Process	$d \log P_t / d \log \sigma_t^2$
1950–84	
AR(1)	−0.014 (0.001)
AR(12)	−0.020 (0.006)
IMA(1,3)	−0.175 (0.038)
1930–84	
AR(1)	−0.022 (0.002)
AR(12)	−0.048 (0.010)
IMA(1,3)	−0.225 (0.026)

Notes: Asymptotic standard errors are shown in parentheses. Each entry shows the percentage change in stock market prices from a 1 percent change in stock market volatility, calculated as described in the text.

small, regardless of the stochastic specification. For the postwar period, the AR(1) process implies an elasticity of share prices with respect to volatility of −.014, while the AR(12) specification implies −.020. Both models suggest that volatility shocks have very small effects on share prices. The moving average specification for volatility differences raises the elasticity to −.175, but even this is an economically small effect. For the full sample period, all of the estimates are somewhat larger. In the AR(1) and AR(12) cases, the elasticities are −.022 and −.048, while in the IMA(1,3) case, the elasticity rises to −.225. A 50 percent increase in market volatility, even using this extreme estimate of volatility persistence, would therefore depress share prices by only 11 percent. These results cast serious doubt on the view that changes in volatility, through their influence on investors' risk premia, have a substantial effect on stock market values.

One potential difficulty with our results is that they describe the persistence of *measured* market volatility, which is a noisy observation on true volatility in any month. It is possible, although we believe unlikely, that market participants have information which enables them to determine the true level of volatility. In this case, the measurement error in our data will bias our estimates toward understating true persistence.

It may be of some interest to analyze the implied persistence in the "true" volatility series. Assuming that the measurement error is independent and identically distributed over time with a zero mean and variance σ_v^2, and that true volatility follows an AR(1) process, then our estimate of ρ_1 will have a probability limit of $\theta\rho_1$, where $\theta = \mathrm{Var}(\sigma_t^2)/[\mathrm{Var}(\sigma_t^2)+\sigma_v^2]$. We estimate θ, the ratio of signal variance to total variance in measured volatility, to be .816 for the postwar period and .890 for the 1930–84 period.[10] These estimates imply that the monthly first-order autocorrelation coefficient for the true volatility process is .700 for the postwar period and .820 for the full sample. Thus, even if market participants could observe true volatility, volatility shocks would have a small effect on market prices.

The general finding that emerges from this section is that stock market volatility is not highly persistent. The data provide little support for the hypothesis that changes in volatility could have an important effect on the level of stock prices. How, then, does Pindyck reach an opposite conclusion using similar data? First, he constructs and graphs a moving average of monthly volatility estimates. Because averaging induces serial correlation, these estimates exhibit substantial persistence. Second, he regresses excess stock returns on the change in his moving average volatility measure, finds a statistically significant effect, and concludes that volatility shocks played an important part in depressing the stock market.

Pindyck's regression of excess returns on changes in volatility does not demonstrate that autonomous volatility changes have had an important effect on market prices. First, since changes in his volatility measure can be predicted, his regression implies that investors should be able to realize substantial profits by trading on past volatility information. Second, the results are contaminated by a serious simultaneity problem. A decline in the stock market should increase volatility through the leverage effect discussed by Black (1976) and Andrew Christie (1982). Finding that movements in the market are negatively correlated with the volatility measure by no means demonstrates the causal role of volatility shocks in depressing the market. Even if volatility changes had no effect on stock prices, one would expect a negative correlation between returns and volatility changes. Third, the implied relationship between volatility shocks and stock returns is weak. French et al. find that the elasticity of share prices with respect to *unanticipated* volatility shocks is at most -0.0272.[11]

---

[10] We calculate the noise-to-signal ratio relying on the fact that the sampling variance of the estimated variance from a normal population is $2\sigma^4/N$, where N is the sample size and σ^4 is the square of the true variance. Using $2\sigma^4/k_t$ as an estimate of each month's measurement error variance, we compute the fraction of the total time-series variation in market volatility which should be accounted for by these errors and use this to calculate θ.

III. Estimates of Serial Correlation in Volatilities Implied by Option Premia

Although the estimates presented in the previous section suggest that volatility shocks are short-lived, they might fail to reflect market participants' beliefs about volatility persistence. The previous estimates also provide little guidance concerning how far back into history investors look in judging the persistence of volatility. Since the volatility process estimated for the postwar period displays substantially less persistence than that for the full sample period, this issue is of some consequence for analyzing how volatility shocks affect share prices.

To address these problems, we analyze the persistence of changes in *ex ante* market volatilities as inferred from option premia. These data represent a market estimate of the *ex ante* volatility which theory says should affect required returns and stock prices. Unfortunately, options on stock market indices such as the Standard and Poor's 500 have been traded for too short a period to make analyzing them informative. However, the Chicago Board Options Exchange (CBOE) has computed an index of the price of a standardized stock option on every Thursday since January 8, 1976:

> ...the CBOE Call Option Index is an average of percent option premiums; for each CBOE underlying stock, a market premium is estimated for a hypothetical six-month, at the money option using the market premiums of existing option series. This estimated market premium is expressed as a percentage of the stock price. The CBOE Call Option Index for a given day is the arithmetic average of all such percent premiums on CBOE underlying stocks on that day. [CBOE, 1979, p. 1]

These data may be used to analyze the persistence of volatility expectations.

The CBOE Index does not correspond to the option premium of any traded security. It is a measure of the option premium on the "representative share" for which options are traded on the CBOE. The implied volatility should therefore be substantially higher than the volatility of the market, since the market is a weighted average of many imperfectly correlated shares. While our estimates of the implied volatility on a representative share are not directly comparable to the volatilities estimated in the last section, their serial correlation properties should be similar.[12] This is supported by our finding that movements in the implied volatilities cohere reasonably well with those of *ex post* volatilities estimated from daily returns on the S&P 500.

To estimate the volatility of the "representative stock" implied by the CBOE Index, we assume that the dividend yield on this share equals that on the S&P 500.[13] We follow Black's (1975) suggestion for dividend adjustment and subtract the present value of dividend payments over the life of the option from the price of the stock. We assume that the option on the representative stock is priced according to the formula derived by Black and Myron Scholes (1973) and apply a numerical search algorithm to determine the variance of returns which is consistent with the observed option price, risk-free rate, and market dividend yield.

---

[11] These calculations are based on the estimates reported on page 25 of French et al.

[12] Henry Latané and Richard Rendleman (1976) compute implied standard deviations for a series of options over a 39-week period and discover that these implied standard deviations tend to move together. Richard Schmalensee and Robert Trippi (1978) find similar results. These coincident movements in volatility are the market-wide volatility shifts we hope to capture.

[13] We assumed that dividends were paid as a continuing flow at rate λ per year, where λ is the current yield on the S&P 500.

Table 3 shows the estimated autocorrelogram and partial autocorrelogram for the implied volatility series. These are *weekly* data, and so the estimated first-order autocorrelation (.965) is higher than those in the last section.[14] The partial autocorrelogram once again suggests that an AR(1) model is an appropriate representation for the series. One year after a shock to volatility, expected volatility is predicted to exceed its mean by only $.965^{52} = .16$, or 16 percent, of the initial shock. The corresponding value of the elasticity of share prices with respect to volatility, calculated by rescaling the parameters α, r, and g to correspond to weekly values, is -0.041. This is roughly comparable to our estimates in the last section, and confirms our earlier conclusion that volatility changes do not persist. These data also support our claim that volatility is a stationary series, since we again reject the null hypothesis of a unit root. The test statistic for the nonstationarity hypothesis, maintaining a first-order autoregressive model, was -3.04, while the .05 critical value from Fuller is -2.88.[15]

TABLE 3—AUTOCORRELOGRAM FOR VOLATILITY FORECASTS IMPLIED BY OPTION PREMIA

Lag Length (weeks)	Autocorrelation	Partial Autocorrelation
1	0.965 (0.012)	0.965 (0.012)
2	0.932 (0.017)	−0.057 (0.046)
3	0.895 (0.021)	−0.094 (0.046)
4	0.859 (0.024)	−0.012 (0.046)
5	0.823 (0.027)	−0.061 (0.046)
6	0.789 (0.029)	−0.019 (0.046)
7	0.753 (0.031)	−0.056 (0.046)
8	0.716 (0.033)	−0.036 (0.047)
9	0.685 (0.034)	−0.031 (0.047)
10	0.655 (0.035)	0.017 (0.047)
11	0.632 (0.036)	0.096 (0.047)
12	0.613 (0.037)	0.034 (0.047)
13	0.596 (0.038)	0.039 (0.047)
14	0.574 (0.039)	−0.117 (0.047)
15	0.551 (0.039)	−0.008 (0.047)
16	0.534 (0.040)	0.099 (0.047)
17	0.515 (0.041)	−0.024 (0.047)
18	0.497 (0.041)	0.016 (0.047)
19	0.484 (0.042)	0.057 (0.047)
20	0.472 (0.042)	0.046 (0.048)
21	0.464 (0.042)	0.038 (0.048)
22	0.458 (0.042)	0.038 (0.048)
23	0.450 (0.042)	−0.039 (0.048)
24	0.443 (0.043)	−0.003 (0.048)

Notes: Volatility forecasts are calculated by inverting the Black-Scholes option valuation formula to obtain the volatility implied by CBOE option premia indices. These data span the period 1976:1 to 1984:26, a total of 447 weekly observations. Standard errors are shown in parentheses.

The CBOE Call Option Index data provide some evidence on the beliefs of market participants, but they do not permit us to directly investigate how long-term expectations of volatility respond to changing short-term volatility expectations. A second source of option data can illuminate this issue. Since 1979, Value Line has computed indices of option premia at three- and six-month maturities. The availability of two different maturity option indices provides an opportunity for additional tests of the persistence hypothesis. We inverted these option premia indices using the same procedure as for the CBOE data.[16]

---

[14] Because each weekly observation on the CBOE Index depends on forecasts of volatility for each of the next 26 weeks, two consecutive observations on the implied volatility will have 25 weeks of forecast volatilities in common. This may bias our estimated autocorrelations. We therefore estimated autoregressive models using nonoverlapping data periods, corresponding to every twenty-sixth observation in our data set. The estimated 6-month autocorrelation coefficient from these data is .43, which is only slightly lower than the 6-month autocorrelation implied by our weekly estimates.

[15] We also performed unit root tests with much longer autoregressive polynomials and continued to reject the unit root hypothesis at very high confidence levels.

[16] The Value Line data were available for the period 1980:16 to 1984:26; this constitutes a total of 220 weeks. However, there were 17 weeks of missing data. This precluded calculating the long autocorrelograms which are reported for the other volatility series. How-

The implied volatility for the six-month options is assumed to equal the average of the expected three-month volatilities for the next three months, and the three months following them. This means:

$$(16) \qquad {}_{t,6}\sigma_t^2 = \left({}_{t,3}\sigma_t^2 + {}_{t,3}\sigma_{t+13}^2\right)/2,$$

where ${}_{s,k}\sigma_t^2$ is the volatility expected to prevail, as of time t, over the k months beginning in week s. The assumption in (16) allows us to solve for an estimate of the implied *forward volatility* that is expected to prevail for the three-month period beginning three months from the current week:

$$(17) \qquad {}_{t+13,3}\hat{\sigma}_t^2 = 2\,{}_{t,6}\hat{\sigma}_t^2 - {}_{t,3}\hat{\sigma}_t^2.$$

Regression estimates of the change in this implied forward volatility which occurs when the current three-month "spot" implied volatility changes are

$$(18) \quad {}_{t+13,3}\hat{\sigma}_t^2 - {}_{t+12,3}\hat{\sigma}_{t-1}^2 = \underset{(.0769)}{-.0233} + \underset{(0.050)}{0.511}[{}_{t,3}\hat{\sigma}_t^2 - {}_{t-1,3}\hat{\sigma}_{t-1}^2].$$

These results indicate that when current volatility expectations change, expected volatility in future periods also changes, but by much less than the change in current volatility. They imply that the half-life of a volatility shock is just over three months, and provide further evidence for our contention that volatility shocks do not persist.

IV. Conclusion and Implications

Our findings suggest that shocks to stock market volatility do not persist for long periods. Estimates based on both actual and *ex ante* volatilities indicate that these volatility shocks have half-lives of less than six months, and in some cases as short as one month. Moreover, our results suggest that substantially greater half-lives for volatility shocks can be rejected at extremely high confidence intervals. Our highest estimate of volatility persistence, obtained by fitting an IMA(1, 3) process to volatilities for the 1930–84 period, suggests that the elasticity of the market price with respect to a volatility shock is −.23. Most of our estimates suggest that this elasticity is much smaller, between −.02 and −.05. These estimates imply that a doubling of volatility, which is a large shock by historical standards, would therefore reduce the level of share prices by at most 23 percent and probably by much less.

These results cast doubt on the hypothesis advanced by Malkiel and Pindyck that the transition to periods of higher uncertainty was responsible for the very poor performance of the stock market during the 1970's. They lead us to doubt that volatility fluctuations, and the movements in equity risk premia which they induce, can explain a large fraction of the variation in the stock market's level. More generally, these results only deepen the puzzle of accounting for the fundamental factors which explain the dramatic fluctuations in stock prices.

---

ever, the first-order autocorrelations for these series, .88 for the 3-month implied volatility and .87 for the 6-month, were roughly consistent with our earlier findings.

REFERENCES

Black, Fischer, "Fact and Fantasy in the Use of Options," *Financial Analysts Journal*, July-August 1975, *31*, 36–41, 61–72.

——, "Studies of Stock Price Volatility Changes," in *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section*, Washington: American Statistical Association, 1976, 177–181.

—— **and Scholes, Myron,** "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, May-June 1973, 81, 637–659.

Brainard, William, Shoven, John and Weiss, Laurence, "The Financial Valuation of the Return to Capital," *Brookings Papers on Economic Activity*, 2: 1980, 453–502.

Christie, Andrew, "The Stochastic Behavior of Common Stock Variances: Value Leverage, and Interest Rate Effects," *Journal of*

Financial Economics, December 1982, *10*, 407–32.

Dickey, David and Fuller, Wayne, "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, July 1981, 49, 1057–1072.

French, Kenneth, Schwert, G. William and Stambaugh, Robert, "Expected Stock Returns and Volatility," Working Paper MERC 85–10, Graduate School of Management, University of Rochester, 1985.

Fuller, Wayne, *The Statistical Analysis of Time Series*, New York: Wiley & Sons, 1976.

Ibbotson, Roger G., *Stocks, Bonds, Bills, and Inflation: 1984 Yearbook*, Chicago: R. G. Ibbotson Associates, 1984.

Kleidon, Allan, "Variance Bounds Tests and Stock Price Valuation Models," Research Paper 806, Stanford University Graduate School of Business, 1985.

Latané, Henry and Rendleman, Richard, Jr., "Standard Deviations of Stock Price Ratios Implied in Option Prices," *Journal of Finance*, May 1976, *31*, 369–81.

Malkiel, Burton, G., "The Capital Formation Problem in the United States," *Journal of Finance*, May 1979, 34, 291–306.

Mankiw, N. Gregory, Romer, David and Shapiro, Matthew, "An Unbiased Reexamination of Stock Market Volatility," *Journal of Finance*, May 1985, *40*, 677–87.

Marsh, Terry A. and Merton, Robert C., "Dividend Variability and Variance Bounds Tests for the Rationality of Stock Market Prices," *American Economic Review*, June 1986, *76*, 483–98.

Merton, Robert C., "An Intertemporal Capital Asset Pricing Model," *Econometrica*, September 1973, *41*, 867–87.

______, "On Estimating the Expected Return on the Market," *Journal of Financial Economics*, December 1980, *8*, 323–61.

Officer, Robert R., "The Variability of the Market Factor of the New York Stock Exchange," *Journal of Business*, July 1973, *46*, 434–53.

Pindyck, Robert S., "Risk, Inflation, and the Stock Market," *American Economic Review*, June 1984, *74*, 335–51.

Schmalensee, Richard and Trippi, Robert, "Common Stock Volatility Expectations Implied by Option Premia," *Journal of Finance*, March 1978, *33*, 129–47.

Schwert, G. William, "Effects of Model Specification on Tests for Unit Roots," Working Paper GPB 85-10, Economics Department, University of Rochester, 1985.

Shiller, Robert, "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?," *American Economic Review*, June 1981, *71*, 421–36.

Chicago Board Options Exchange, *The CBOE Call Option Index: Methodology and Technical Considerations*, Chicago, 1979.

The Cyclical Flexibility of Wages

By CHARLES L. SCHULTZE*

In a recent issue of this *Review*, Anthony O'Brien (1985) takes issue with the findings of my 1981 article that the short-run flexibility of wages in response to cyclical changes in aggregate demand was already small by the early years of this century and did not decline substantially after World War II.

In the 1981 paper, I constructed several measures of wage and price flexibility. One of them was a coefficient of cyclical flexibility for various indices of prices and wages derived as the ratio between two variables: 1) the mean change in inflation rates in expansions less their mean change in contraction phases of the business cycle; and 2) the mean change in nominal *GNP* growth in expansions less their mean change in contraction phases. That ratio I called the flexibility coefficient, λ, and compared its values for early peacetime cycles, lying in the interval 1900–14, 1923–29, with its values for the first part of the postwar period, 1949–66.[1] Among other results I found that for the private nonfarm deflator and the nonfood *CPI*, flexibility was already quite low in the prewar years and did not decline into the postwar period. With farm and food prices included, the flexibility coefficients for the total *GNP* deflator and the total *CPI*, while of only moderate size in the prewar period fell even lower in the postwar years. In the case of average wages for the economy as a whole, flexibility was also low in the prewar period ($\lambda = 0.16$),[2] and it became somewhat lower in the postwar years ($\lambda = 0.10$ to 0.11). In the case of manufacturing wages to which O'Brien pays particular attention, flexibility, while still not substantial, was somewhat higher in both periods (0.25 and 0.16 respectively).[3] The flexibility coefficients for wholesale prices did fall substantially. As I concluded in the 1981 paper, "Wage flexibility was never large; but the combination of its smaller size and the lower amplitude of contractions in the postwar period produced a wage series that rose only slightly less in contractions than expansions" (p. 541).

O'Brien addresses my findings with respect to wages and criticizes both the choice of wage series and the substantive analysis of the article. I shall not discuss the data issues here, except to note that with respect to manufacturing wages—where O'Brien concentrates his analytical comments—my earlier results are essentially unchanged when one substitutes the prewar series he prefers (the Rees manufacturing wage series) for the one I used (Rees through 1919 spliced into a BLS wage series thereafter). For the calculations in the remainder of this note, I use the Rees series over the entire period, 1890–1940.

In his analytical comments, O'Brien raises some important issues, but their resolution does not, I believe, substantially alter the conclusions of my paper. O'Brien argues that my exclusion of the very large post-World War I cycle of 1919–21, the 1890–99 years which include a depression in 1892–94, and the Great Depression years, 1930–40, substantially biases downward my prewar coefficient of manufacturing wage flexibility. In fact, however, the addition of both the depression years 1930–40 and the decade of the 1890's to my data lowers my previously published coefficient of manufacturing wage

---

*Senior Fellow, The Brookings Institution, 1775 Massachusetts Avenue, NW, Washington, D.C. 20036.

[1] The 1981 article explained the reasons for this choice of periods.

[2] In my earlier article, the economywide wage flexibility coefficient for the prewar period was mislabeled "private nonfarm" and given a value of 0.15. It should have been labeled "total economy" and given a value of 0.16. The postwar wage flexibility of 0.10 to 0.11 does refer to the private nonfarm economy.

[3] Using perfect market clearing as a standard ($\lambda = 1.0$), the decline in flexibility between the two periods should be judged by comparing the values of $(1 - \lambda)$. Thus, manufacturing wages fell 75 percent short of perfect flexibility in the peacetime cycles of the prewar period and 84 percent short in the postwar years.

flexibility from 0.25 to 0.20, and takes it a little closer to the postwar value for manufacturing of 0.16. Adding in the 1930's when manufacturing wages first declined to 1933, rose very substantially to 1937 and then fell in 1938 raises the flexibility coefficient. But adding in the 1890's, when manufacturing wages accelerated during the year after the cyclical peaks of 1892 and 1895, lowers the coefficient. O'Brien added back only the depression years, and so obtained a larger coefficient of 0.35.

For purposes of comparing pre- and postwar wage flexibility in recessions, the depression of the 1930's should, I believe, be excluded. In terms of the concepts I used in the 1981 article, the huge changes in unemployment and the large swings in prices that occurred during the 1929–40 period led to shifts in the preexisting wage and price norms. More importantly, insofar as manufacturing wages are concerned, the onetime wage gains in the years after 1933, from New Deal measures and unionization, substantially affected the results.[4] For both these reasons, the cyclical swings in wages relative to the swings in nominal *GNP* as measured by λ were greater in the 1929–40 period than in the typical peacetime business cycle.

Why not include the immediate postwar cycles, 1919–21 and 1946–49, in the sample of data from which λ is estimated? The 1981 paper has a section (pp. 562–63) which shows that the behavior of wages and prices in the war and immediate postwar cycles, for both World War I and World War II, is vastly different from their behavior at other times. I argue in that paper that economic agents recognized the atypical nature of the wartime expansions and the accompanying wartime inflation. And so norm or expected rates of inflation collapsed very rapidly in the postwar readjustment recessions relative to their behavior in other recessions. In the swing from expansion in 1919–20 to contraction in 1920–21, the ratio of the deceleration of manufacturing wages to the deceleration of nominal *GNP* (i.e., the λ) was 0.80; for the post-World War II adjustment cycle (1946–49) the ratio was also high, 0.61. I do not believe these atypical responses should be included when analyzing the peacetime cyclical behavior of inflation. But, even when I do include both postwar adjustment cycles, along with the recovery of 1921–23, in estimating λ, its value becomes larger but is still very much the same in both eras: 0.30 for the early period (1890–1914; 1919–40) and 0.29 for the later period (1946–66).

In sum, there are good analytical reasons for excluding the two postwar adjustment cycles and the depression of the 1930's from my sample. But their inclusion, along with the cycles of the 1890's, does not essentially alter the conclusions I drew in the 1981 paper about the comparative behavior of manufacturing wage inflation: It was already quite sticky in the late nineteenth and early twentieth centuries, but became somewhat stickier in the years after World War II.

REFERENCES

Baily, Martin Neil, "The Labor Market in the 1930s," in James Tobin, ed., *Macroeconomics, Prices and Quantities*, Washington: The Brookings Institution, 1983.

O'Brien, Anthony, "The Cyclical Sensitivity of Wages," *American Economic Review*, December 1985, *75*, 1124–32.

Rees, Albert, *Real Wages in Manufacturing, 1890–1914*, NBER, Princeton: Princeton University Press, 1961.

______, *New Measures of Wage-Earner Compensation in Manufacturing, 1914–57*, NBER Occasional Paper 75, New York, 1960.

Schultze, Charles L., "Some Macroeconomics Foundations for Micro Theory," *Brookings Papers on Economic Activity*, 2:1981, 521–92.

---

[4] See, for example, the discussion by Martin Neil Baily (1983, pp. 53–56).

A New View of the Federal Debt and Budget Deficits: Comment

By Ross D. Milbourne and Daniel J. Richards*

In a recent paper (1984), Robert Eisner and Paul Pieper, present new calculations of the real value of the federal debt. Eisner and Pieper correctly take account of fluctuations in the market value of nominal assets and liabilities, and also take account of changes in real values wrought by inflation. They define an adjusted real deficit as the change in the real net debt of the government (liabilities less assets). Defining $L(A)$ as the nominal value of already outstanding government liabilities (assets), P as the price level and $\dot{P}$ as its rate of change, and D as the conventionally measured deficit, Eisner-Pieper define their adjusted real deficit (D^*/P) as

$$(1)\quad \frac{D^*}{P} \equiv \frac{D}{P} + \frac{(\Delta L - \Delta A)}{P} - \frac{(L-A)}{P}\dot{P}.$$

Eisner and Pieper not only claim that it is a new measure, but a better measure of fiscal policy, and provide econometric evidence to that effect.

Eisner and Pieper interpret D^*/P as a measure of "fiscal thrust," the *ex post actual* effect of spending and taxation upon income. However, they also adjust the *high-employment* deficit, a measure not of actual spending but of "conscious fiscal action."[1] The Eisner-Pieper measure, however, is not better in separating endogenous and discretionary factors.

A central difficulty with the Eisner-Pieper deficit is that the price and interest rate effects it includes are the intended result of policy, not policy *actions* themselves. Restrictive fiscal policy will lower both price levels and interest rates, which may be what originally motivated the policy. But both effects raise the real market value of outstanding debt, which Eisner and Pieper consider expansionary. Thus, the Eisner-Pieper deficit could label a deliberate and successful contraction as expansionary.[2] They argue that if expansionary policy causes inflation and leads to no change in real variables, it cannot be labelled expansionary. Analogously, consider an increase in the money supply which raises prices and so does not change the real money supply in the long run. Eisner and Pieper argue that, since when the resulting price changes are taken into account no real variables are affected, the original increase in the money supply is not expansionary.

Eisner and Pieper claim that their deficit measure is better, on the basis of regressions of *GNP* growth (ΔY) on various deficit measures (d) as follows:

$$(2)\qquad \Delta Y_t = \alpha + \beta d_{t-1} + \gamma d_{t-2}.$$

Their deficit measure produces better coefficients and a higher R^2. But this *must* happen, given the totally inappropriate reduced form specification (2). To see this, consider the simple macro model

$$Y = C + I + G$$

$$C = c(Y-T) + h(L/P), \quad I = \bar{I},\ T = \bar{T},$$

---

*Department of Economics, Queen's University, Kingston, Canada K7L 3N6, and Department of Economics, Tufts University, Medford, MA. 02155, respectively. We thank Doug Purvis, Robert Eisner, and Paul Pieper for comments.

[1] See Walter Heller (1967, pp. 66–67).

[2] The 1937–39 recession is a case in point. Cary Brown (1956) has shown this recession was in part due to contractionary fiscal policy (the conventional high-employment surplus being $1.4 billion in 1937). But back-of-the-envelope calculations show that the Eisner-Pieper high-employment deficit for those years would classify fiscal policy as *expansionary*, due to a contemporary price decline of 5 percent on federal debt of $29 billion.

where C, I, G, and T are, consumption, investment, government spending, and lump sum taxes, respectively, all in real terms. Lump sum taxation is used to remove a distinction between the actual and high-employment deficit, and real wealth is assumed to depend upon the real value of bonds or government liabilities (L/P). Defining $D/P = G - T$, the reduced form of this model gives, when differenced.

$$(3) \qquad \Delta Y_t \equiv Y_t - Y_{t-1} = \frac{\Delta\bar{I} + c\Delta\bar{T}}{1-c} + \frac{1}{1-c}\left(\left(\frac{D}{P}\right)_t - \left(\frac{D}{P}\right)_{t-1}\right) + \frac{h}{1-c}\Delta\left(\frac{L}{P}\right)_t.$$

Clearly, real wealth effects will enter the reduced form,[3] so that in running a simple, misspecified regression such as (2) which excludes real wealth (Pigou) effects, a deficit measure which incorporates those wealth effects, (D^*/P), must do better. Eisner and Pieper's results merely show that wealth effects were important, which confirms Frederic Miskin's (1977) results. Eisner and Pieper have simply obtained a higher R^2 by transferring the Pigou effect to the fiscal policy variable.

While Eisner-Pieper's figures are a better indication of the government's *ex post* position, they are not a better measure of fiscal policy.

REFERENCES

Brown, E. Cary, "Fiscal Policy in the 'Thirties': A Reappraisal," *American Economic Review*, December 1956, *46*, 857–79.

Eisner, Robert and Pieper, Paul J., "A New View of the Federal Debt and Budget Deficits," *American Economic Review*, March 1984, *74*, 11–29.

Heller, Walter, *New Dimensions of Political Economy*, New York: W. W. Norton, 1967.

Mishkin, Frederic, "What Depressed the Consumer: The Household Balance Sheet and the 1973–75 Recession," *Brookings Papers on Economic Activity*, 1:1977, 123–64.

---

[3] Of course, assuming that the first term in (3) is a constant over the sample period, as done in (2), is also misspecifying the reduced form.

A New View of the Federal Debt and Budget Deficits: Reply

By ROBERT EISNER AND PAUL J. PIEPER*

It is not clear what Ross Milbourne and Daniel Richards mean in their comment (1986) by referring to the official high-employment deficit as a measure of "conscious fiscal action," as they add immediately that our adjusted measure (1984) "is not better in separating endogenous and discretionary factors" (p. 1154). In fact, the high-employment deficit is *not* a measure of conscious fiscal action. It is simply the deficit that would exist at a specified level of employment. That may reflect, in part, conscious decisions by fiscal authorities. It also reflects much more.

As we pointed out originally, "The high-employment budget, while aimed at eliminating the cylical effects on the actual budget, is of course not exogenous to what is going on in the economic system. To consider it a measure of fiscal thrust is not to view it as entirely at the discretion of fiscal authorities." In particular, we noted that "higher price levels and inflation cause more than proportionate increases in corporate taxes, and in personal taxes as long as the rate structure is progressive and not indexed" (pp. 17–18). Expenditure programs are similarly affected by inflation and changing interest rates, both of which may be quite outside the realm of "conscious fiscal action." The Fed, for example, can increase the actual and (unadjusted) high-employment budget deficits by taking action that increases nominal interest rates.

Our adjusted measures of deficits do remove some of the nondiscretionary elements in the official, unadjusted high-employment budget by in effect adjusting for the impact of inflation on nominal interest payments. But their essential merit is rather that, for discretionary, nondiscretionary, and endogenous elements, they come closer to showing the real deficit, that is, the real increase in the government's net debt.

As for Milbourne and Richards' concern that our measure might mischaracterize a contractionary budget as expansionary, we first must distinguish time periods. Thus, an increase in taxes of $50 billion is likely to reduce the current high-employment deficit by $50 billion dollars. This would be true for both the official high-employment deficit and our adjusted one. If the consequence of this is lower prices and interest rates, there will be an ultimate impact on federal expenditures and revenues. If these lower prices and interest rates are assumed in estimating subsequent high-employment deficits, those subsequent deficits—both official and adjusted—will be affected. Whether a lower nominal deficit will so reduce prices and interest rates as to increase the subsequent real deficit is extremely doubtful, but the price and interest effects on both nominal and real deficits should certainly be recognized.[1]

But since the lower prices and interest rates will come about only to the extent that the tax increase is contractionary, it is difficult to lend credibility to Milbourne and Richards' argument that our measure could "label a deliberate and successful contraction as expansionary." If the tax increase lowers interest rates and prices, it is presum-

---

*William R. Kenan Professor of Economics, Northwestern University, Evanston, IL 60201 and Assistant Professor of Economics, University of Illinois at Chicago, Chicago, IL 60680, respectively.

[1] Milbourne and Richards' interpretation of fiscal policy in the 1937–38 recession does not square with the facts. The *GNP* implicit price deflator rose 4.6 percent from 1936 to 1937, and between December 1936 and December 1937 consumer prices rose by 3.2 percent (U.S. Department of Commerce, *National Income and Product Accounts of the United States, 1929–74,* p. 359, and *Statistical Abstract of the United States,* 1940, p. 332). The inclusion of price effects on the real value of debt would have made the high-employment surplus larger, not turned it into a deficit. The 1937–38 downturn was thus all the better explained by the adjusted budget. Prices did decline in 1938, and this then may well have contributed, along indeed with the breakout of war in Europe, to the 1939 recovery.

ably because it lowers aggregate demand. The lower interest rates and lower prices then offer a *partial* offset to the contractionary force of the tax increase. If the interest and price effects are so great as to increase the adjusted deficit, then real aggregate demand would actually be higher, so interest rates and prices could not be lower.

Milbourne and Richards' argument is similar to the claim that a decrease in the quantity of money could lower interest rates, because the higher interest rates it would bring on would so reduce income as to lower the demand for money more than the supply of money. One can construct dynamic expectations scenarios with such perverse reactions, but they appear highly implausible. Where perverse effects do occur, however, we would be better off to recognize them, as our adjusted measure would.

There is a further analogy to monetary policy which may be instructive. An increase in the rate of growth of the nominal money supply will not prove expansionary in real terms if it is accompanied by an equal increase in inflation: the real rate of growth of money supply will be unchanged. Similarly, a greater nominal deficit, increasing the rate of growth of the nominal debt, will not be expansionary if an increased rate of inflation prevents the real debt from growing any faster. It may well be argued, though, that in both cases inflation would prevent real changes only if the economy were already at full employment.

As to when policies, fiscal or monetary, would be "labelled expansionary" (Milbourne and Richards' words, not ours), if they increase prices and nominal income they are clearly expansionary in nominal terms. If Milbourne and Richards wish to say that, we have no quarrel. And just as clearly, if policies do not change real variables, they are not expansionary in real terms.

Commenting on our empirical results, Milbourne and Richards point out that our adjusted measures perform better because they include real wealth changes. That is exactly the point. But they argue that wealth changes should not be included in a measure of fiscal policy. Wealth changes are, however, in the unadjusted deficit. Adapting Milbourne and Richards' notation, they are only the nominal changes, $\Delta L/P$. We reject this implicit money illusion and include the real changes, $\Delta(L/P)$. We have thus treated the inflation tax analogously to other taxes: both result in transfers of wealth from the private sector to the government.

REFERENCES

Eisner, Robert and Pieper, Paul J., "A New View of the Federal Debt and Budget Deficits," *American Economic Review*, March 1984, *74*, 11–29.

Milbourne, Ross D. and Richards, Daniel F., "A New View of the Federal Debt and Budget Deficits: Comment," *American Economic Review*, December 1986, *76*, 1154–55.

Government Debt, Government Spending, and Private Sector Behavior: Comment

By JAMES R. BARTH, GEORGE IDEN, AND FRANK S. RUSSEK*

In a recent issue of this *Review* (1983), Roger Kormendi relied upon the work of Martin Bailey (1962, 1971) and Robert Barro (1974) "to develop an alternative approach to modeling private sector consumption-saving behavior based on rational evaluation of the consequences of government fiscal policy" (p. 995). Briefly, this approach involves determining to what extent government spending substitutes for private consumption, and to what extent government debt is net wealth and thereby also affects private consumption. This "consolidated approach" to examining private sector consumption is not only novel, but appears at a time when there is widespread interest as well as controversy over the effects of government fiscal policy. It therefore seems only appropriate to examine further the results obtained by Kormendi—not to criticize them but rather to explore their robustness.

This comment specifically explores the sensitivity of Kormendi's results to alternative time periods of estimation, to the differentiation between federal debt and state and local government debt, and to the measurement of government debt—particularly par vs. market value.

I. Empirical Model

To test the consolidated approach, Kormendi specifies and estimates the following two equations:

$$(1) \quad PC_t = a_0 + a_{11}Y_t + a_{12}Y_{t-1} + a_2 GS_t + a_3 W_t + a_4 TR_t + a_5 GB_t + u_t,$$

$$(2) \quad PC_t = a_0 + a_{11}Y_t + a_{12}Y_{t-1} + a_2 GS_t + a_3 W_t + a_4 TR_t + a_5 GB_t + a_6 TX_t + a_7 RE_t + a_8 GINT_t + u_t,$$

where both equations are in first-difference form and PC is private sector consumption, Y is net national product, W is private wealth, GS is government spending on goods and services, TR is transfer payments, TX is taxes, RE is corporate retained earnings, $GINT$ is government interest payments on its outstanding debt, and GB is government debt. As regards equation (1), the consolidated approach implies that the coefficient a_2 should be negative and between zero and minus one, and that the coefficient a_5 should be zero, whereas the "standard approach" assumes the private sector ignores government spending so that a_2 is zero and treats government debt as net wealth so that a_5 is positive. As regards equation (2), the consolidated approach implies that $-1 < a_2 < 0$, and that a_5 through a_8 are all zero.[1] The

---

*Barth: George Washington University, Washington, D.C. 20052; Iden and Russek: Congressional Budget Office, Washington, D.C. 20515. This paper was written when Barth was on leave at the Congressional Budget Office. We are grateful to Roger Kormendi for his helpful suggestions, and to Stacy A. Miller and Peter A. Woodward for their excellent research assistance. The views expressed in this paper are our own and do not necessarily reflect the views of the Congressional Budget Office or its staff.

[1] Kormendi states that "The magnitude of a_4 will then depend upon the discrepancies in propensities to consume and the extent to which changes in transfers are perceived to be permanent" (p. 997). Holding government purchases constant, however, transfer payments usually are treated as negative taxes and thus as having no effect on consumption under the Ricardian Equivalence Theorem (see for example, Willem Buiter and James Tobin, 1980, and Martin Feldstein, 1982). As the referee noted, "Treating transfer payments as negative taxes with no wealth redistribution is *less* general than treating them as potentially redistributive." But he goes on to state, "Of course, taxes may redistribute wealth, which is a potential reason for TX having a zero coefficient."

TABLE 1—KORMENDI'S ORIGINAL RESULTS AND OUR ATTEMPTED REPLICATIONS[a]

Independent Variable[b]	1931–76 (1)	Replication (2)	1931–76 (3)	Replication (4)	1931–40/ 1947–76 (5)	Replication (6)
Intercept	N.A.	0.003	N.A.	0.005	N.A.	0.005
		(0.81)		(1.23)		(1.10)
Y	0.29	0.32	0.32	0.32	0.33	0.33
	(7.3)	(8.21)	(14.8)	(14.27)	(14.1)	(13.67)
Y_{-1}	0.07	0.07	0.06	0.06	0.05	0.06
	(3.3)	(3.00)	(3.0)	(2.87)	(2.2)	(2.61)
GS	−0.23	−0.21	−0.23	−0.21	−0.21	−0.20
	(12.8)	(11.05)	(13.2)	(11.10)	(3.5)	(3.22)
W	0.03	0.02	0.03	0.02	0.03	0.02
	(3.0)	(3.06)	(3.4)	(3.51)	(3.5)	(2.95)
TR	0.83	0.64	0.78	0.63	0.74	0.64
	(5.6)	(3.78)	(5.6)	(3.94)	(3.5)	(3.80)
TX	0.07	0.02	–	–	–	–
	(0.9)	(0.20)				
RE	0.10	−0.01	–	–	–	–
	(0.9)	(0.12)				
$GINT$	1.15	1.67	–	–	–	–
	(1.3)	(1.88)				
GB	−0.06	−0.05	−0.04	−0.03	−0.03	−0.04
	(2.9)	(2.76)	(2.4)	(1.91)	(1.6)	(2.02)
$\bar{R}^2$	0.911	0.880	0.902	.875	0.910	0.881
SER	0.0175	0.0178	0.0176	0.0181	0.0178	0.0185

[a] Dependent variable-consumer expenditures; first-difference form. The t-statistics are shown in parentheses.

[b] Y = net national product, constant dollars, per capita; GS = federal, state, and local government purchases of goods and services, *NIPA* bais, constant dollars, per capita; W = private wealth, constant dollar per capita, measured by the stock of residential plus nonresidential fixed capital plus an estimate of the stock of human capital; TR = federal, state, and local government transfers to persons, constant dollars, per capita; TX = federal, state, and local government revenues, constant dollars, per capita; RE = corporate retained earnings, constant dollars, per capita; $GINT$ = federal, state, and local government net interest payments, constant dollars, per capita; GB = federal, state, and local government debt, market value, beginning of period, constant dollars, per capita; GBP = par value counterpart of GB; GBF = federal government component of GB; $GBSL$ = state and local government component of GB; $GBFP$ = federal government component of GBP; and $GBSLP$ = state and local government component of GBP.

standard approach, on the other hand, implies that a_2 is zero, that a_4, a_5, and a_8 are positive, and that a_6 and a_7 are negative.

Table 1 presents Kormendi's results of estimating equations (1) and (2) as well as our attempted replications.[2] In addition to estimating the two equations for the entire 1931–76 period, Kormendi estimated equation (1) with the war years omitted. As Table 1 shows, our attempted replication of Kormendi's results leaves his general conclusions unaltered. Government spending significantly substitutes for private consumption and government debt fails to be significantly and positively related to consumption.[3] Furthermore, although government interest payments are nearly significant, taxes

---

[2] The Data Appendix documents all of the data used in this comment. Although we were not able to obtain his original data, we wish to thank Kormendi for helping us in our reconstruction of his data set. Also, we are grateful to James Butkiewics, W. Michael Cox, John Musgrave, and John Seater for data and helpful discussions.

[3] Also, our finding that government debt and private wealth have significantly different coefficients is generally inconsistent with the standard approach.

TABLE 2—ANALYSIS OF KORMENDI'S RESULTS: EFFECTS OF UPDATING THROUGH 1983 AND SEPARATING FEDERAL FROM STATE AND LOCAL GOVERNMENT DEBT[a,b]

Independent Variable	1931–83 (7)	1931–83 (8)	1931–40/ 1947–83 (9)	1931–83 (10)	1931–83 (11)	1931–40/ 1947–83 (12)
Intercept	0.01	0.01	0.01	0.01	0.01	0.02
	(1.82)	(2.44)	(2.51)	(1.64)	(2.63)	(3.28)
Y	0.31	0.31	0.31	0.31	0.32	0.32
	(6.77)	(11.99)	(11.24)	(6.60)	(12.07)	(12.13)
Y_{-1}	0.06	0.05	0.05	0.06	0.04	0.05
	(2.43)	(2.02)	(1.90)	(2.39)	(1.92)	(1.91)
GS	−0.20	−0.19	−0.17	−0.20	−0.20	−0.17
	(9.21)	(9.04)	(2.41)	(8.89)	(9.18)	(2.48)
W	0.02	0.02	0.02	0.02	0.02	0.01
	(2.98)	(2.93)	(2.52)	(2.95)	(2.75)	(2.24)
TR	0.69	0.72	0.71	0.69	0.67	0.59
	(3.85)	(4.21)	(3.96)	(3.78)	(3.87)	(3.38)
TX	−0.02			−0.02		
	(0.21)			(0.22)		
RE	0.07			0.07		
	(0.55)			(0.57)		
$GINT$	1.42			1.50		
	(2.62)			(2.28)		
GB	−0.04	−0.02	−0.01			
	(1.97)	(1.09)	(0.62)			
GBF				−0.04	−0.01	0.01
				(1.88)	(0.86)	(0.48)
$GBSL$				−0.02	−0.09	−0.16
				(0.31)	(1.57)	(2.49)
$\overline{R}^2$	0.828	0.812	0.816	0.824	0.815	0.836
SER	0.0213	0.0223	0.0229	0.0215	0.0221	0.0217

[a,b]See Table 1.

and retained earnings clearly have no significant effect on private consumption. Based upon an F-test, one cannot reject the null hypothesis of the consolidated approach that $a_6 = a_7 = a_8 = 0$. Thus, Kormendi's original results and our attempted replications provide some support for the consolidated as opposed to the standard approach to explaining consumer behavior. Viewing government transfer payments as negative taxes, however, one can reject the null hypothesis that $a_4 = a_6 = a_7 = a_8 = 0$. In addition, the finding that government debt has a significantly negative effect on consumption appears inconsistent with both approaches.[4]

The next step in our check on the robustness of Kormendi's findings was to update the analysis through 1983. The results of this exercise are presented in columns 7 through 9 in Table 2. As may be seen, these results are broadly similar to those reported in Table 1. However, government interest payments are now significant, while government debt retains its negative sign but becomes insignificant at the 5 percent level. This latter finding is consistent with the consolidated approach. Moreover, based upon an F-test, one still cannot reject the null hypothesis of the consolidated approach that $a_6 = a_7 = a_8 = 0$. However, this hypothesis can be rejected based upon Wald and likelihood ratio tests at the 5 percent level.[5] For all these

---

[4] In his referee's comment, Kormendi noted that "A negative coefficient on government debt is *not* inconsistent with the Consolidated Approach when adjusted to include uncertainty of tax incidence." In this regard, see pp. 1004–05 in Kormendi's original paper.

[5] According to the referee, "Under normal error, F-test is exact, while others are asymptotic." However, G. B. A. Evans and N. E. Savin (1982, p. 740) have

TABLE 3—ANALYSIS OF KORMENDI'S RESULTS: POSTWAR PERIODS, EFFECTS OF UPDATING AND SEPARATING FEDERAL FROM STATE AND LOCAL GOVERNMENT DEBT[a,b]

Independent Variable	1947–76 (13)	1947–76 (14)	1947–76 (15)	1947–76 (16)	1947–83 (17)	1947–83 (18)	1947–83 (19)	1947–83 (20)
Intercept	0.02	0.02	0.03	0.03	0.03	0.03	0.04	0.04
	(2.06)	(1.90)	(3.56)	(3.75)	(3.86)	(4.38)	(5.09)	(6.09)
Y	0.20	0.30	0.21	0.31	0.19	0.24	0.19	0.26
	(2.37)	(7.21)	(2.81)	(8.94)	(2.15)	(6.36)	(2.42)	(8.09)
Y_{-1}	0.06	0.05	0.09	0.06	0.05	0.03	0.07	0.04
	(1.39)	(1.65)	(2.32)	(2.14)	(1.15)	(0.89)	(1.97)	(1.49)
GS	−0.13	−0.17	−0.18	−0.17	−0.14	−0.14	−0.17	−0.15
	(1.60)	(2.12)	(2.45)	(2.52)	(1.97)	(1.90)	(2.57)	(2.45)
W	0.01	0.02	0.004	0.01	0.02	0.02	0.01	0.01
	(1.14)	(1.41)	(0.42)	(0.75)	(2.28)	(2.11)	(0.76)	(1.47)
TR	0.61	0.58	0.44	0.37	0.39	0.41	0.28	0.23
	(2.20)	(2.15)	(1.79)	(1.59)	(1.73)	(1.85)	(1.33)	(1.17)
TX	0.14		0.19		0.10		0.14	
	(0.98)		(1.55)		(0.68)		(1.13)	
RE	0.11		0.22		0.12		0.23	
	(0.57)		(1.30)		(0.68)		(1.36)	
$GINT$	1.96		−0.63		1.15		−0.72	
	(1.69)		(0.47)		(2.09)		(0.87)	
GB	−0.03	−0.02			0.01	0.03		
	(1.01)	(0.73)			(0.50)	(0.96)		
GBF			0.02	0.02			0.06	0.05
			(0.52)	(0.62)			(1.91)	(2.18)
$GBSL$			−0.32	−0.28			−0.27	−0.19
			(3.02)	(3.52)			(2.56)	(2.95)
$\bar{R}^2$	0.761	0.744	0.822	0.826	0.744	0.721	0.795	0.800
SER	0.0196	0.0203	0.0169	0.0167	0.0207	0.216	0.0185	0.0183

[a,b]See Table 1.

tests, moreover, one can reject the hypothesis that $a_4 = a_6 = a_7 = a_8 = 0$.

Table 2 also presents results of estimating equations (1) and (2) with government debt broken down into its federal and state and local components. These results are reported in columns 10 through 12. As may be seen, the government spending variable remains highly significant and carries a negative coefficient that lies between zero and minus one. The federal debt variable is insignificant, though it does become positive in one case. The state and local government debt variable is negative in all three equations and quite significant in the one omitting the war years. Based upon an F-test, one can reject $a_{GBF} = a_{GBSL}$ in equation (12) but not in (10) or (11). Also, $a_6 = a_7 = a_8 = 0$ cannot be rejected.

The next set of results is presented in Table 3 and covers the shorter, more recent periods of 1947 through 1976 and 1947 through 1983. Columns 13 through 16 contain the results for the first period (which is the postwar part of Kormendi's sample period), while columns 17 through 20 contain the results for the second period (which updates the postwar part of Kormendi's sample period). Results for both equations (1) and (2) are presented as well as for both total government debt and its separate federal and state and local components. Again, $a_6 = a_7 = a_8 = 0$ cannot be rejected, but the results for the government spending variable are sensitive to omitting the prewar time period; the coefficient becomes insignificant (at the 5 percent level) for equation (2) in the 1947–76 period and for equations (1) and (2) in the 1947–83 period. As

---

shown that, in the normal error case, these tests are monotonic functions of the F-test statistic, so that there should be no conflict among them. Thus, our findings suggest that the normality assumption may be questionable and its potential violation less than innocuous.

Table 4—Analysis of Kormendi's Results: Effects of Altering Postwar Sample Periods[a,b]

Independent Variable	1947–81 (21)	1947–81 (22)	1947–81 (23)	1947–81 (24)	1949–83 (25)	1949–83 (26)	1949–83 (27)	1949–83 (28)	1949–82 (29)	1955–81 (30)
Intercept	0.02	0.02	0.03	0.03	0.02	0.03	0.04	0.03	0.03	0.03
	(2.08)	(2.30)	(3.56)	(3.62)	(2.74)	(3.67)	(5.08)	(5.40)	(4.51)	(3.50)
Y	0.20	0.30	0.20	0.30	0.21	0.24	0.24	0.27	0.23	0.31
	(2.61)	(7.84)	(2.95)	(8.95)	(2.35)	(6.21)	(3.36)	(9.18)	(3.32)	(4.18)
Y_{-1}	0.07	0.05	0.09	0.60	0.07	0.04	0.14	0.10	0.14	0.09
	(1.77)	(1.65)	(2.61)	(2.14)	(1.59)	(0.91)	(3.60)	(3.06)	(3.56)	(2.02)
GS	−0.18	−0.18	−0.20	−0.19	−0.10	−0.12	−0.10	−0.08	−0.12	−0.17
	(2.68)	(2.70)	(3.34)	(3.11)	(1.33)	(1.51)	(1.61)	(1.37)	(1.79)	(1.63)
W	0.02	0.02	0.01	0.01	0.02	0.02	−0.0001	0.01	0.001	0.003
	(2.76)	(2.66)	(1.20)	(2.01)	(2.16)	(2.03)	(0.01)	(1.01)	(0.19)	(0.36)
TR	0.51	0.53	0.40	0.35	0.50	0.44	0.46	0.41	0.45	0.50
	(2.44)	(2.63)	(2.09)	(1.87)	(2.00)	(1.71)	(2.38)	(2.15)	(2.31)	(2.21)
TX	0.16		0.20		0.04		0.06		0.08	0.09
	(1.21)		(1.69)		(0.29)		(0.56)		(0.65)	(0.80)
RE	0.14		0.24		0.11		0.23		0.23	0.03
	(0.85)		(1.56)		(0.59)		(1.58)		(1.54)	(0.19)
GINT	0.68		−0.99		1.39		−0.97		−0.92	−1.77
	(1.27)		(1.30)		(2.35)		(1.32)		(1.23)	(2.38)
GB	−0.02	−0.01			0.04	0.03				
	(0.68)	(0.52)			(0.95)	(0.91)				
GBF			0.02	0.01			0.12	0.12	0.10	0.13
			(0.75)	(0.54)			(3.44)	(3.61)	(2.41)	(2.49)
GBSL			−0.28	−0.18			−0.34	−0.22	−0.34	−0.37
			(2.92)	(2.95)			(3.57)	(3.76)	(3.51)	(3.94)
$\bar{R}^2$	0.799	0.791	0.842	0.837	0.687	0.647	0.809	0.804	0.812	0.842
SER	0.0187	0.0191	0.0166	0.0169	0.0208	0.0221	0.0163	0.0165	0.0164	0.0151

[a,b] See Table 1.

regards the total government debt variable, it becomes positive in the 1947–83 period, but is not significant in either period. When the debt variable is partitioned, the federal component is uniformly positive and becomes significant in equation (1) for the 1947–83 period. The state and local component, on the other hand, is uniformly negative and significant for both periods, and the *F*-test statistics indicate that $a_{GBF} = a_{GBSL}$ can be rejected in all cases.[6] Overall, these results provide less support for the consolidated approach than do the original results reported in Table 1.

Table 4 presents still further results for selected postwar sample periods. To determine to what extent the results were dependent upon the inclusion of 1982 and 1983, these two years were omitted. Columns 21 through 24 therefore present the results for the period 1947 through 1981. The government spending variable is negative and significant in all four cases, while the total government debt variable is negative but insignificant. When the debt variable is partitioned into its federal and state and local components, the federal component is positive but insignificant, while the state and local component is negative and significant. Again, $a_{GBF} = a_{GBSL}$ can be rejected in all cases, but $a_6 = a_7 = a_8 = 0$ cannot be rejected.[7]

To be symmetric, the first two years were dropped from the sample period rather than the last two years.[8] The results of this particular estimation are reported in columns 25 through 28 of Table 4. In this case, the government spending variable is still negative but no longer significant. The total government debt variable now becomes positive, but is not significant. However, when this latter variable is partitioned into its federal and state and local components, the federal component becomes positive and significant, while the state and local component remains negative and significant.[9]

---

[6] According to the referee, "*GBF* positive and *GBSL* negative are contradictory." He adds that "no theory predicts such results." However, Buiter and Tobin state, "The claim that households internalize public debt is surely more credible for state and local governments. They are subject to legal and economic debt limits; they cannot print money; there are fairly direct links between their outlays, including debt service, and their tax levies" (pp. 46–47). Together with Kormendi's uncertainty argument, the Buiter-Tobin view suggests a negative coefficient on state and local debt is not inconsistent with federal debt having a positive coefficient.

[7] These findings hold throughout for Table 4.

[8] Coincidentally, this means that we begin with the same year as that used by Buiter and Tobin.

To further check on the robustness of the findings for the postwar period, we examined the sensitivity of the results for the 1949 through 1983 period to the exclusion of 1983, as well as results for the much shorter but more recent 1955–81 period.[10] Our findings for these periods indicate that the government spending variable is negative but insignificant, that the federal debt variable is positive and significant, and that the state and local debt variable is negative and significant.[11] These findings demonstrate that the previous results questioning the consolidated approach are not entirely due to the inclusion of the years 1982 and 1983. More important, they further demonstrate that the choice of sample period can indeed affect the empirical results one obtains. This point, of course, applies more generally than to just the Kormendi study.

The next-to-last check on Kormendi's results involved using the par as opposed to market value of government debt. The reason for making this particular modification is that many previous researchers have relied upon the former measure, and therefore it would be useful to determine whether Kormendi's results are sensitive to this substitution. The results of estimating equations (1) and (2) for various periods based upon a par value measure of the government debt variable are presented in Table 5. Overall these results show that the government spending variable is always negative, but only significant for the longer periods of 1931–81 and 1931–83. For the shorter postwar periods of 1949–83, 1949–82, and 1952–82, this variable is never significant.[12] On the other hand, the government debt variable when measured on a par value basis is negative for the longer periods and positive for the postwar periods. Interesting enough, it is consistently significant only in equation (2) for the former periods and in equation (1) for the shorter and more recent periods.[13] As far as the consolidated approach is concerned, these results offer mixed support.

As one final check, both the market value of government debt and the difference between the market and par values of government debt were included together in the same regression equations.[14] These results —which parallel those reported in Table 5 —are presented in Table 6. Note that the difference between the market and par values of government debt is a measure of the capital gain, and that this term when subtracted from the market value of government debt yields the par value of government debt. (Recall that all variables are in first-difference form.) The results in Tables 5 and 6 should therefore be consistent with one another. For the most part, this is the case. Whenever the par value of government debt is insignificant in Table 5, the two debt variables included in Table 6 are also insignificant, with one exception (compare col. 39 in Table 5 and col. 49 in Table 6). Whenever the par value of government debt is significant in Table 5, both debt variables in Table 6 are also significant, except in the longer periods when the market value of government debt is significant but the difference between it and its par value is not.

---

[9] In his referee comment, Kormendi observed, "The fact that *TX* is always small and insignificant suggests that deficits don't matter. This does not square with the stock of debt mattering." However, even if federal debt is net wealth, "taxes may redistribute wealth, which is a potential reason for *TX* having a zero coefficient" (see fn. 1).

[10] Beginning in 1955 corresponds to the starting year for some of the empirical results on government debt and interest rates reported in our earlier article (1984).

[11] When equation (1) was reestimated for the 1949–82 and 1955–81 periods, the same results were obtained except that the coefficient on federal debt was not quite significant at the 5 percent level in the latter period. In these equations, the *F*-test statistics indicated rejection of $a_{GBF} = a_{GBSL}$, but not rejection of $a_6 = a_7 = a_8 = 0$.

[12] Starting in 1952 corresponds to the year Levis Kochin (1974) used to begin his study.

[13] Based on the *F*-test statistics, we can reject $a_6 = a_7 = a_8 = 0$ for 1931–83, but not for 1931–81 or the postwar periods.

[14] The referee suggested that this specification be explored and reported.

TABLE 5—ANALYSIS OF KORMENDI'S RESULTS: PAR VALUE OF GOVERNMENT DEBT[a,b]

Independent Variable	1931–81 (31)	1931–81 (32)	1931–83 (33)	1931–83 (34)	1949–83 (35)	1949–83 (36)	1949–82 (37)	1949–82 (38)	1952–82 (39)	1952–82 (40)
Intercept	0.003	0.005	0.008	0.01	0.03	0.03	0.02	0.03	0.02	0.03
	(0.84)	(1.39)	(1.63)	(2.39)	(3.27)	(4.23)	(2.77)	(3.73)	(2.74)	(3.55)
Y	0.31	0.32	0.31	0.31	0.19	0.19	0.20	0.20	0.22	0.21
	(8.19)	(14.71)	(6.92)	(12.04)	(2.32)	(5.17)	(2.32)	(5.06)	(2.48)	(5.17)
Y_{-1}	0.06	0.05	0.07	0.05	0.07	0.04	0.07	0.05	0.05	0.04
	(2.91)	(2.51)	(2.69)	(1.97)	(1.63)	(1.23)	(1.66)	(1.27)	(1.01)	(1.13)
GS	−0.21	−0.20	−0.20	−0.19	−0.04	−0.02	−0.06	−0.04	−0.04	−0.02
	(11.48)	(11.71)	(9.33)	(9.07)	(0.47)	(0.29)	(0.73)	(0.47)	(0.40)	(0.21)
W	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	(4.73)	(5.02)	(3.22)	(3.27)	(2.64)	(2.85)	(2.73)	(2.87)	(2.42)	(2.49)
TR	0.65	0.65	0.71	0.72	0.50	0.46	0.48	0.44	0.51	0.51
	(4.34)	(4.54)	(4.02)	(4.16)	(2.19)	(2.03)	(2.06)	(1.93)	(1.79)	(1.86)
TX	0.03		−0.02		0.01		0.03		0.009	
	(0.41)		(0.19)		(0.08)		(0.20)		(0.06)	
RE	−0.01		0.05		0.08		0.07		−0.04	
	(0.06)		(0.41)		(0.46)		(0.43)		(0.21)	
$GINT$	1.03		1.69		0.97		1.02		0.83	
	(1.97)		(2.97)		(1.70)		(1.76)		(1.35)	
GBP	−0.04	−0.02	−0.05	−0.02	0.14	0.17	0.10	0.15	0.13	0.17
	(2.40)	(1.43)	(2.33)	(1.02)	(2.18)	(3.01)	(1.41)	(2.21)	(1.58)	(2.26)
$\bar{R}^2$	0.884	0.879	0.833	0.811	0.727	0.725	0.730	0.726	0.720	0.730
SER	0.0178	0.0181	0.0209	0.0223	0.0194	0.0195	0.0196	0.0198	0.0202	0.0198

[a,b]See Table 1.

TABLE 6—ANALYSIS OF KORMENDI'S RESULTS: MARKET AND MARKET MINUS PAR VALUE OF GOVERNMENT DEBT[a,b]

Independent Variable	1931–81 (41)	1931–81 (42)	1931–83 (43)	1931–83 (44)	1949–83 (45)	1949–83 (46)	1949–82 (47)	1949–82 (48)	1952–82 (49)	1952–82 (50)
Intercept	0.004	0.006	0.01	0.01	0.03	0.03	0.02	0.03	0.03	0.03
	(0.94)	(1.54)	(1.56)	(2.39)	(3.47)	(4.53)	(3.00)	(3.94)	(3.10)	(3.70)
Y	0.31	0.32	0.31	0.31	0.19	0.19	0.19	0.21	0.23	0.23
	(8.19)	(15.20)	(6.85)	(11.82)	(2.27)	(5.34)	(2.29)	(5.44)	(2.69)	(6.13)
Y_{-1}	0.06	0.05	0.06	0.05	0.08	0.05	0.08	0.05	0.06	0.06
	(3.08)	(2.87)	(2.61)	(1.98)	(1.78)	(1.42)	(1.86)	(1.56)	(1.39)	(1.82)
GS	−0.21	−0.21	−0.19	−0.19	−0.03	−0.01	−0.60	−0.05	−0.06	−0.06
	(11.40)	(11.90)	(8.87)	(8.60)	(0.33)	(0.16)	(0.68)	(0.55)	(0.59)	(0.63)
W	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01
	(4.28)	(4.42)	(3.07)	(2.90)	(1.89)	(2.09)	(2.04)	(2.22)	(1.24)	(1.38)
TR	0.66	0.64	0.70	0.72	0.55	0.50	0.53	0.49	0.71	0.72
	(4.40)	(4.60)	(3.95)	(4.13)	(2.36)	(2.28)	(2.26)	(2.20)	(2.47)	(2.77)
TX	0.04		−0.02		0.01		0.03		−0.004	
	(0.54)		(0.20)		(0.04)		(0.19)		(0.03)	
RE	0.01		0.05		0.12		0.12		0.001	
	(0.13)		(0.36)		(0.69)		(0.70)		(0.01)	
$GINT$	0.82		1.72		0.62		0.62		0.11	
	(1.50)		(2.90)		(0.95)		(0.96)		(0.15)	
GB	−0.04	−0.03	−0.05	−0.02	0.16	0.20	0.13	0.16	0.19	0.20
	(2.46)	(1.81)	(2.30)	(1.06)	(2.44)	(3.39)	(1.67)	(2.42)	(2.25)	(2.91)
$GB\text{-}GBP$	−0.006	−0.04	0.05	0.003	−0.22	−0.27	−0.20	−0.25	−0.31	−0.32
	(0.14)	(1.07)	(1.02)	(0.08)	(2.23)	(3.35)	(1.91)	(2.99)	(2.51)	(3.62)
$\bar{R}^2$	0.885	0.885	0.830	0.808	0.730	0.741	0.737	0.750	0.750	0.783
SER	0.0177	0.0177	0.0212	0.0225	0.0194	0.0189	0.0194	0.0189	0.0190	0.0178

[a,b] See Table 1.

Overall, the results in Table 6 do not fully support the consolidated approach. Government spending has a negative but not consistently significant coefficient.[15] The market value of government debt and the difference between it and the par value of government debt are neither uniformly negative and significant nor uniformly insignificant.[16] However, even for the postwar period, the results do not fully support the standard approach either, in particular because the capital gain variable is not positive and significant.[17]

II. Conclusions

This comment has raised but not resolved questions concerning the impact of government fiscal actions on private sector consumption. Kormendi presented new and important empirical results bearing on the issue of whether government spending substitutes for private spending and whether government debt is net worth and thus a stimulant to private consumption. Clearly, his results merit careful scrutiny given the considerable importance of the issues he addresses. It is for this reason that we have updated his original results through 1983, analyzed separately the effects of federal and state and local government debts on private consumption, checked the sensitivity of his results by replacing the market value of government debt with the par value, and concentrated on more recent postwar time periods. Overall, the results based upon these modifications raise sufficient questions about the robustness and interpretation of Kormendi's original findings that more empirical work in this important research area is clearly needed.

DATA APPENDIX

This Data Appendix corresponds to that presented in Kormendi (1983, pp. 1007–08). Its purpose is to document the updating of data and the construction of additional variables appearing in this comment.

The following abbreviations are used: *Banking and Monetary Statistics, 1914–1940* (*BMS1*); *Banking and Monetary Statistics, 1941–1970* (*BMS2*); *Current Population Reports* (*CPR*); *Flow of Funds Accounts, Assets and Liabilities Outstanding, 1959–1982* (*FF*); *Historical Statistics of the United States, Colonial Times to 1970* (*HS*); *National Income and Product Accounts of the United States, 1929–1976* (*NIPA*); *Survey of Current Business* (*SCB*); *Standard and Poor's Statistical Service* (*SPSS*); *Standard and Poor's Statistical Service, Current Statistics* (*SPSSCS*).

The primary data series are

POP = Total Resident Population 1929–83 (*CPR*, Series P-25).

PNNP = Implicit Price Deflator for Net National Product 1929–83 (Table 7.5, line 3 in *NIPA*, and in *SCB*, July 1982, July 1983, and March 1984 issues).

NNPR = Net National Product in billions of 1972 dollars 1930–83 (Table 1.8, line 3 in *NIPA*, and in *SCB*, July 1982, July 1983, and March 1984 issues).

CNDR = Personal Consumption Expenditures on Nondurable Goods in billions of 1972 dollars 1930–83 (Table 1.2, line 4 in *NIPA*, and in *SCB*, July 1982, July 1983, and March 1984 issues).

CSR = Personal Consumption Expenditures on Services in billions of 1972 dollars 1930–83 (Table 1.2, line 5 in *NIPA*, and in *SCB*, July 1982, July 1983, and March 1984 issues).

CDR = Personal Consumption Expenditures on Durable Goods in billions of 1972 dollars 1930–83 (Table 1.2, line 3 in *NIPA*, and in *SCB*, July 1982, July 1983, and March 1984 issues).

GSR = Government Purchases of Goods and Services in billions of 1972 dollars 1930–83 (Table 1.2, line 21 in *NIPA*, and in *SCB*, July 1982, July 1983, and March 1984 issues).

TXN = Total Government Receipts in billions of current dollars 1930–83 (Table 3.1, line 1 in *NIPA*, and in *SCB*, July 1982 and July 1983 issues; and Table 3.2, line 1 plus Table 3.3, line 1 in *SCB*, March 1984).

GSN = Total Government Expenditures in billions of current dollars 1930–83 (Table 3.1, line 6 in *NIPA*, and in *SCB*, July 1982 and July 1983 issues; and Table 3.2, line 12 plus Table 3.3, line 13 in *SCB*, March 1984).

TRN = Total Government Transfer Payments to Persons in billions of current dollars 1930–83 (Table 3.1, line 11 in *NIPA*, and in *SCB*, July 1982 and July 1983 issues; and Table 3.2, line 17 plus Table 3.3, line 17 in *SCB*, March 1984).

GINTN = Total Government Interest Payments to Persons and Business less Dividends Received by Government in billions of current dollars 1930–83 (Table 3.1, line 13 minus line 16 minus line 18 in *NIPA*, and in *SCB*, July 1982 and July 1983 issues; and Table 3.2, line 20 minus line 23 plus Table 3.3, line 18 minus line 21 in *SCB*, March 1984).

---

[15] Moreover, $a_6 = a_7 = a_8 = 0$ can be rejected for the 1931–83 sample period.

[16] See John Seater (1981) for tests of the Ricardian Equivalence Theorem that incorporate both of these debt-related variables. Based on data through 1974, he does not find consistent support for the debt-neutrality view, in part because the market value of government debt and the capital gain both have significantly positive effects on some components of private consumption.

[17] Perhaps, however, Kormendi's uncertainty argument applies here (see fn. 4 above). For additional empirical results, see our article (1986).

REN = Undistributed Corporate Profits (Retained Earnings) with Inventory Valuation and Capital Consumption Adjustments in billions of current dollars 1930–83 (Table 5.1, line 4 in *NIPA*, and in *SCB*, July 1982, July 1983, and March 1984 issues).

$DURR$ = Total Net Stock of Durable Goods Owned by Consumers in billions of 1972 dollars beginning of period 1930–83 (*SCB*, March 1979, p. 24, Table 4, first col. of data; revisions for 1966–78 and updates for 1979–83 were obtained directly from John Musgrave of the Bureau of Economic Analysis).

$WN1$ = Total Private National Wealth in Current Prices in billions of dollars 1930–46 beginning of period (Raymond Goldsmith, 1956, pp. 14–15, Table W-1, col. 1 minus col. 9 minus col. 17 minus col. 24).

$WN2$ = Total Private National Wealth in Current Prices in billions of dollars 1947–59 beginning of period (*HS*, p. 255, Series F422 minus series F429 minus series F437 minus series F444).

$WN3$ = Total Private Wealth in Current Prices in billions of dollars 1960–69 beginning of period (*HS*, p. 252, series F349 minus series F352 minus series F361 minus series F364).

KN = Total Net Stock of Fixed Nonresidential plus Residential Private Capital in billions of current dollars 1930–78 beginning of period (*SCB*, February 1981, p. 58, Table 2, first col. of data plus p. 62, Table 6, second col. of data; updates for 1979–83 were obtained directly from John Musgrave, BEA).

HN = Net Stock of Human Capital in Education and Training in billions of current dollars 1930–69 beginning of year (Kendrick, 1976, pp. 204–05, Table B-20, fourth col. of table).

$GBN1$ = Market Value of Privately Held Total Government Debt in billions of dollars beginning of period 1930–76 (Seater, 1981, pp. 90–91, Table 1, col. labelled *MVTOTOG*1).

$GBN2$ = Market Value of Privately Held Marketable Federal Government Debt in billions of dollars beginning of period 1977–81 (W. Michael Cox and Eric Hirschhorn, 1983, p. 268, Table 6, last col. of data).

$GBN3$ = Market Value of State and Local Government Debt in billions of dollars beginning of period 1930–76 (Seater, 1981, pp. 92–93, Table 2, col. labelled *MVSL*).

$GBFP1$ = Par Value of Privately Held Federal Government Debt in billions of dollars beginning of period 1930–40 (*BMS1*, p. 512, Table 149, col. 5, lagged average of June numbers through 1935, lagged December numbers thereafter).

$GBFP2$ = Par Value of Privately Held Federal Government Debt in billions of dollars beginning of period 1941–61 (*BMS2*, p. 882, Table 13.4, col. 5, lagged December numbers).

$GBFP3$ = Par Value of Privately Held Federal Government Debt in billions of dollars beginning of period 1962–83 (*FF*, p. 2, line 3 minus p. 16, line 9).

$GBFPSB$ = Par Value of U.S. Savings Bonds in billions of dollars beginning of period 1977–83 (*FF*, p. 14, line 17).

$GBSLP$ = Par Value of State and Local Government Debt in billions of dollars beginning of period 1977–83 (*FF*, p. 12, line 14).

$MPRA$ = Ratio of Market Value to Par Value of Outstanding Public Debt beginning of period 1977–81 (James Butkiewicz, 1983, pp. 376–79, Appendix table, col. labelled Ratio 1, December values from preceeding year; for 1982–83, the values of $MPRA$ were obtained directly from Butkiewicz).

MBP = Municipal Bond Price per dollar beginning of period 1930–82 (*SPSS*, 1982, p. 233, Table entitled Municipal Bond Price Indexes, December values from preceding year; the December 1982 value of MBP was obtained from *SPSSCS*, January 1983, p. 35, last row of col. 15).

The following transformations of the above data were used to generate the variables used in our replication, updating, and modifications of the Kormendi regressions. The presentation closely resembles that in Kormendi (p. 1008).

1) $PC=(CNDR+CSR+.3\times DURR+.1\times CDR)/POP$

2) $Y=NNPR/POP$

3) $GS=GSR/POP$

4) $TR=TRN/(PNNP\times POP)$

5) $TX=TXN/(PNNP\times POP)$

6) $RE=REN/(PNNP\times POP)$

7) $GINT=GINTN/(PNNP\times POP)$

8) $W=WN/(.5\times PNNP\times POP+.5\times PNNP_{-1}\times POP_{-1})$

where

$WN=WN1+HN$ (1930–46);
$=WN2+HN$ (1947–59);
$=WN3(WN2_{1959}/WN3_{1959})+HN$ (1960–69);
$=KN(WN_{1969}/KN_{1969})$ (1970–83).

9) $GB=GBN/(.5\times PNNP\times POP+.5\times PNNP_{-1}\times POP_{-1})$

where

$GBN=GBN1$ (1930–76);
$=GBN2+GBFPSB\times MPRA+GBSLP\times MBP$ (1977–81);
$=GBFP3\times MPRA+GBSLP\times MBP$ (1982–83).

10) $GBF=GBFN/(.5\times PNNP\times POP+.5\times PNNP_{-1}\times POP_{-1})$

where

$GBFN=GBN1-GBN3$ (1930–76);
$=GBN2+GBFPSB\times MPRA$ (1977–81);
$=GBFP3\times MPRA$ (1982–83).

11) $GBSL=GBSLN/(.5\times PNNP\times POP+.5\times PNNP_{-1}\times POP_{-1})$

where

$GBSLN=GBN3$ (1930–76);
$=GBSLP\times MBP$ (1977–83).

12) $GBP=GBPN/(.5\times PNNP\times POP+.5\times PNNP_{-1}\times POP_{-1})$

where

$GBPN=GBFP1+GBN3/MBP$ (1930–40);
$=GBFP2+GBN3/MBP$ (1941–61);
$=GBFP3+GBN3/MBP$ (1962–76);
$=GBFP3+GBSLP$ (1977–83).

REFERENCES

Bailey, Martin J., *National Income and the Price Level*, New York: McGraw-Hill, 1962, 1971.

Barro, Robert J., "Are Government Bonds Net Wealth?," *Journal of Political Economy*, November/December 1974, *82*, 1095–1117.

Barth, James R., Iden, George and Russek, Frank S., Jr., "Do Federal Deficits Really Matter?," *Contemporary Policy Issues*, Fall 1984–85, *3*, 79–95.

______, ______, and ______, "The Economic Consequences of Federal Deficits: An Explanation of the Net Wealth and Instability Issues," *Southern Economic Journal*, July 1986, *53*, 27–50.

Buiter, Willem H. and Tobin, James, "Debt Neutrality: A Brief Review of Doctrine and Evidence," in George M. von Furstenberg, ed., *Social Security versus Private Saving*, Ballinger, 1980, 39–63.

Butkiewicz, James L., "The Market Value of Outstanding Government Debt, Comment," *Journal of Monetary Economics*, May 1983, *11*, 373–80.

Cox, W. Michael and Hirschhorn, Eric, "The Market Value of U.S. Government Debt: Monthly, 1942-1980," *Journal of Monetary Economics*, March 1983, *11*, 261–72.

Evans, G. B. A. and Savin, N. E., "Conflict Among the Criteria Revisited: The *W*, *LR*, and *LM* Tests," *Econometrica*, May 1982, *50*, 737–48.

Feldstein, Martin, "Government Deficits and Aggregate Demand," *Journal of Monetary Economics*, January 1982, *9*, 1–20.

Goldsmith, Raymond W., *A Study of Savings in the United States*, Vol. III, Princeton: Princeton University Press, 1956.

Kendrick, John W., *The Formation and Stocks of Total Capital*, New York: National Bureau of Economic Research, 1976.

Kochin, Levis A., "Are Future Taxes Anticipated by Consumers?," *Journal of Money, Credit and Banking*, August 1974, *6*, 385–94.

Kormendi, Roger C., "Government Debt, Government Spending, and Private Sector Behavior," *American Economic Review*, December 1983, *73*, 994–1010.

Seater, John J., "The Market Value of Outstanding Government Debt, 1919-1975," *Journal of Monetary Economics*, July 1981, *8*, 85–102.

______, "Are Future Taxes Discounted?," *Journal of Money, Credit and Banking*, August 1982, *14*, 376–89.

Government Debt, Government Spending and Private Sector Behavior: Comment

By FRANCO MODIGLIANI AND ARLIE STERLING*

Roger Kormendi (1983) presents apparently strong evidence that, in contrast to the standard view, consumption *is not* reduced by taxes but *is* reduced by government expenditure. He interprets his results as supporting a "consolidated" approach to private sector behavior in which consumers effectively internalize the government budget constraint. Specifically, he claims that consumers regard government spending as the true measure of the government's claim on private resources, and so do not respond to changes in taxes, given spending. This is basically the approach advocated by Robert Barro (1974) and known also as the Ricardian Equivalence Proposition (hereafter REP).

Kormendi's results appears to contradict other empirical work based on the Life Cycle Hypothesis (for example, Martin Feldstein, 1982; Modigliani 1984a; Sterling, 1985) although results similar to his have been reported (see David Aschauer, 1985; John Seater and Roberto Mariano, 1985 and the references in Kormendi). In our view, Kormendi's analysis is seriously flawed. His heuristic derivation of the consumption function leads him to a specification of the aggregate consumption function, which is not consistent with the Life Cycle Hypothesis (LCH) or with REP, and to questionable methods of estimation. Once his conceptual and methodological errors are corrected, his formulation and, more generally, the REP hypothesis are found to receive little empirical support.

In the next section we rely on the LCH to derive an aggregate consumption function which shows explicitly how government expenditure and taxes should effect private consumption. This derivation helps to bring out the observable implications of REP, which are shown to be equivalent to a limiting form of the LCH in which the planning horizon is infinite. It also serves to clarify the appropriate specification of the variables appearing in the consumption function. Next, Section II reports our empirical estimates and tests. Section III compares our results with Kormendi's. Finally, Section IV reports the results of endeavors to improve the specification of fiscal variables, notably by distinguishing between permanent and transitory tax changes.

I. Government Spending and Taxes in the Aggregate Consumption Function

We begin with the standard model of aggregate consumption, in which consumption, C_t, is proportional to the sum of beginning-of-period aggregate nonhuman wealth, A_t, and aggregate human wealth:

(1)
$$C_t = c\left[A_t + \sum_{i=1}^{R} m_i(Y_{t+i} - T_{t+i})(1+\mu)^{-i}\right],$$

where $\sum_{i=1}^{R} m_i(Y_{t+i} - T_{t+i})(1+\mu)^{-i}$ equals the weighted present value of anticipated after-tax labor income and R is the length of life. The m_i weights equal the proportion of consumers present at time t who are expected to be alive at time $t+i$; Y_{t+i} represents before-tax labor income and T_{t+i} taxes on labor income. After-tax income is discounted at the rate μ,and c is the propensity to consume. The specification (1) is derived in Sterling. It can be shown to hold even if one allows for the presence of a bequest motive. Olivier Blanchard (1985) analyzes a

---

*Sloan School of Management, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02139, and Marsoft Incorporated, One Kendall Square, Cambridge, MA 02139, respectively. James Meehan provided capable research assistance.

similar model in a general equilibrium setting.

Equation (1) seems to support the standard view that consumption depends only on taxes. But the relevant taxes are *future* as well as current ones. As a result, one can readily show, by taking into account the intertemporal government budget constraint that, in general, C will depend on both taxes and government expenditure. The government budget constraint can be written as

$$T_t = E_t + rG_{t-1} - D_t, \tag{2}$$

where E_t represents government expenditures on goods and services, rG_{t-1} interest payments on beginning-of-period government debt, and D_t the current deficit. Using (2), the present value of weighted discounted taxes can be expressed in terms of government expenditures on goods and services and the deficit:

$$\begin{aligned}
&\sum_{i=1}^{R} m_i T_{t+i}(1+\mu)^{-i} \\
&\quad = \sum_{i=1}^{R} m_i E_{t+i}(1+\mu)^{-i} \\
&\qquad + rG_t \sum_{i=1}^{R} m_i (1+\mu)^{-i} \\
&\qquad + r \sum_{i=1}^{R} \sum_{j=1}^{i-1} D_{t+j} m_i (1+\mu)^{-i} \\
&\qquad - \sum_{i=1}^{R} m_i D_{t+i}(1+\mu)^{-i} \\
&\quad = \sum_{i=1}^{R} m_i E_{t+i}(1+\mu)^{-i} + s_1 G_t \\
&\qquad - \sum_{i=1}^{R} m_i (1-b_i) D_{t+i}(1+\mu)^{-i}
\end{aligned} \tag{3}$$

where we have used the identity $G_{t+i} = G_{t-1} + \Sigma_{j=1}^{i} D_{t+j}$ and define

$$s_i = r \sum_{j=i}^{R} m_j (1+\mu)^{-j}; \tag{4}$$

$$b_i = s_{i+1}/m_i(1+\mu)^{-i}.$$

The coefficients s_i measure the present value to those living at time t of the stream of interest paid by a dollar of debt, where the present value is calculated at the personal discount rate μ and the mortality rate is accounted for by the m_i factors. The b_i coefficients measure the same present value, but for those alive at the time the debt is issued.

Substituting for taxes in the consumption function yields

$$\begin{aligned}
C_t = c\Bigg[& A_t - s_1 G_t + \sum_{i=1}^{R} m_i Y_{t+i}(1+\mu)^{-i} \\
& - \sum_{i=1}^{R} m_i E_{t+i}(1+\mu)^{-i} \\
& + \sum_{i=1}^{R} m_i (1-b_i) D_{t+i}(1+\mu)^{-i} \Bigg].
\end{aligned} \tag{5}$$

Or, using (2) and letting T_t^* denote taxes net of all transfers, including interest (i.e., $T^* = E_t - D_t$), one finally obtains

$$\begin{aligned}
C_t &= c\Bigg[A_t - s_1 G_t + \sum_{i=1}^{R} m_i [Y_{t+i} - E_{t+i} \\
&\qquad + (1-b_i)(E_{t+i} - T^*_{t+i})](1+\mu)^{-i} \Bigg] \\
&= c\Bigg[A_t - s_1 G_t + \sum_{i=1}^{R} m_i [Y_{t+i} \\
&\qquad - (1-b_i) T^*_{t+i} - b_i E_{t+1})](1+\mu)^{-i} \Bigg].
\end{aligned} \tag{6}$$

Taxes net of interest appear in (6) because, as noted above, the coefficients s_i and

b_i reflect the anticipated taxes due to an issue of debt. Thus, that component of taxes that is due to interest payments must not be included in the explicit tax term, as it is already incorporated in the effect of the debt.

Equation (6) shows that consumption depends, in general, on a weighted average of net taxes and government expenditure on goods and services, where the weights ($1-b_i$ and b_i, respectively) sum to one—and that it should respond only partially to wealth in the form of government debt. There is only one limiting case in which only expenditure matters while taxes and debt matter not at all, namely, *if* μ *were equal to* r, *and the planning horizon* R *were infinite*. In this case, in fact, $m_i = 1$ for all i, and, as can be seen from (4), s_1 and the b_i would all be unity. In that special case, the consumption function reduces to

$$(7)\quad C_t = c\left[A_t - G_t + \sum_{i=1}^{\infty}(Y_{t+i} - E_{t+i})(1+r)^{-i}\right].$$

It will be recognized that (7), the limiting form of LCH for an infinite planning horizon, embodies the relevant implications of REP; consumption depends on: (*i*) a measure of private net worth which excludes all government debt, and (*ii*) on government spending on goods and services, but is unaffected by the way that expenditure is financed, whether by taxes or by selling debt. Nor is this surprising since, in effect to derive REP, Barro postulates an infinite planning horizon for all economic agents, through the intermediation of an infinite chain of intergenerational bequests.

However, there are many considerations that suggest rejecting the Barro hypothesis of an infinite horizon and recognizing that that representative household will behave as though R is very much finite and not much larger, if at all, than the length of life, as suggested by the standard "life horizon" Life Cycle Hypothesis (LH-LCH). The relevant considerations have been elaborated at length by many authors (see, for example, Modigliani, 1984a,b, and James Tobin, 1980, ch. III). Similarly, there are grounds for hypothesizing that μ may tend to exceed r (Fumio Hayashi, 1982).

These considerations suggest that the coefficient of debt, s_1, and that of expenditure, b_i, in the consumption function should fall well below unity, while that of taxes should not be far from unity. Sterling has shown that, when the population age distribution is approximated by an exponential, s_i, varies between 0.6 (for $r = \mu = 0.02$) and 0.2 (for $r = 0.02$, $\mu = 0.08$) for $i = 1$ and that it declines monotonically with i, as the stream of interest, of which s_i is the weighted present value, gets shorter. These results imply that, under the LH-LCH, the liability for future interest may offset the value of government debt held in private portfolios, but to a modest extent at best. Similarly, for given taxes, an increased deficit, and hence government expenditure, may dampen consumption because of anticipated future taxes to pay interest, but, again, quite moderately at best.

Our conclusions with respect to the role of fiscal variables in the consumption function can now be summarized as follows: (*i*) consumption should respond negatively to a weighted average of taxes and expenditure; (*ii*) the coefficient of the weighted average should equal minus that of income (or, equivalently, the coefficients of the income and that of the weighted average should add up to zero) reflecting the fact that, whether consumption depends on taxes or on government expenditure, it must respond dollar per dollar to the combination of the two; (*iii*) the weight of taxes should be not far from unity, and that of government consumption much less than unity, whereas REP's infinite planning horizon hypothesis would imply the reverse; (*iv*) the coefficient of government debt, given private net worth, should be close to zero (and might even be somewhat positive,[1] whereas, according to REP, it should equal minus that of wealth.

---

[1] If (*i*) government real interest included in T^* is subject to great variability, as typically happens during period of high and variable inflation; (*ii*) consumption

II. Estimates of the Aggregate Consumption Function

The consumption function (6) cannot be estimated directly since it is in terms of unobservable expectations of future income, taxes and expenditure (or deficits). We assume that the expectations can be modeled as a distributed lag of past values of these variables. Accordingly, the basic equation estimated and used in testing alternative hypothesis takes the form

$$(8)\quad C_t = a + b_0 A_t + b_1 G_t + \sum_{i=1}^{L} c_i (Y_{t-i} - T_{t-i}) + \sum_{i=1}^{L} d_i D_{t-i}.$$

The last two terms of (8) are based on a regrouping of the expression in the square bracket in the summation term of (6):

$$Y_{t+1} - (1 - b_i) T^*_{t+i} - b_i E_{t+i} = Y_{t+i} - T^*_{t+i} - b_i D_{t+i},$$

where D_{t+i} denotes the budget deficit, $E_{t+1} - T^*_{t+1}$.

This formulation has the desirable property of insuring that the constraint that the tax and expenditure coefficients add up to minus the income coefficient, as dictated by (6). Since the coefficients d_i in (8) correspond to $-b_i$ in (6), while b_1 in (8) corresponds to s_1 in (6), according to the conventional LCH formulation, both coefficients should be close to zero, whereas according to the REP/Kormendi hypothesis, the Σd_i should be close to $-\Sigma c_i$, while b_1 should be close to $-b_0$.

The explanatory variables appearing in (8) have slightly different definitions from the analogous variables appearing in (6). The changes are necessary to accommodate the expectations formation mechanism and data limitations. Specifically, income and net taxes include income from and taxes on capital, and are measured, respectively, by net national product and all government taxes (net of transfers including government net domestic interest payments). The deficit equals the difference between government expenditures on goods and services and net taxes. Net interest payments to the domestic sector equal the *ex post* real payments, calculated by subtracting the product of the actual rate of inflation and the stock of net government debt from net interest payments.[2,3] In order to insure comparability with Kormendi, all tests reported here are based on annual data.

The c_i and d_i, $i = 1, \ldots, L$ distributed lag coefficients in (8) reflect both the expectations formation mechanism and behavioral parameters. Sterling has shown how the behavioral coefficients can be estimated directly in such a formulation. For the purposes of this comment we shall interpret them as behavioral parameters. In all results reported below, the distributed lag coefficients are constrained to lie on a first-degree polynomial with five lags (including the contemporaneous term). Only the sum of the current and lagged coefficients is reported.

Column 1 of Table 1 reports the results of estimating (8) over the 33 postwar years 1952–84. The results are seen to provide solid support for the LCH and unequivocal rejection of Kormendi/REP, especially in the light of the very close fit (the standard error is 0.47 percent of the mean). First, for REP the sum of the d_i coefficients should be negative and equal in magnitude to the sum of the income coefficients; for LCH it should be close to zero. The actual estimate is not very significantly different from zero and

---

depends positively on the permanent real rate and not current rate (as is true for the LCH); (*iii*) the permanent real rate is approximately constant, say ρ, then (6) might contain an additional term proportional to ρG so that the coefficient of G is the sum of a positive and a negative term, whose sign is uncertain.

[2] See the Data Appendix for a more detailed description of the data.

[3] This approximation is justified by the consideration that for the United States in the postwar period, inflation appears to be fairly well forecastable—see Robert Barsky (1985).

TABLE 1—ESTIMATES OF THE AGGREGATE CONSUMPTION FUNCTION, 1952–84[a,b]

Independent Variables		(1)[c]	(2)[c]	(3)[c]	(4)
Constant	a	−0.38	−0.45	−0.436	−0.436
		(0.06)	(0.12)	(0.052)	(0.061)
Net Worth (A)	b_0	0.023	0.027	0.022	0.022
		(0.005)	(0.005)	(0.004)	(0.005)
Government Debt (G)	b_1	0.073	0.106	0.096	0.096
		(0.024)	(0.033)	(0.021)	(0.031)
Net National Product (Y)	Σc_{1i}	0.922	0.901	0.928	0.931
		(0.020)	(0.041)	(0.017)	(0.070)
Net Taxes (T)	Σc_{2i}	−0.922	−0.901	−0.928	−1.081
		(0.020)	(0.041)	(0.017)	(0.151)
Deficit (D)	Σd_i	0.171	0.115	0.187	–
		(0.082)	(0.086)	(0.069)	
Government Expenditure (E)	Σd_i^*	–	–	–	0.146
					(0.136)
Transfers (TR)	Σt_i	–	0.005	–	–
			(0.102)		
Transitory Taxes (T^*)	Σh_i	–	–	0.633	0.645
				(0.182)	(0.339)
SE		0.015	0.014	0.012	0.013
Log Likelihood		96.33	98.50	103.44	103.88
D-W		1.43	1.73	1.92	1.90

[a] Dependent variable: Consumption per capita; shown in thousands of 1972 dollars. Standard errors are shown in parentheses.

[b] The mean of the dependent variable is 3.21 and its standard deviation is 0.73.

[c] In these equations, the coefficients of taxes, Σc_{2i}, are constrained to equal the negative of the income coefficients, $-\Sigma c_{1i}$.

actually positive. Second, the estimate of b_1, far from being close to $-b_0$, as implied by REP, is seen to be positive, which is not inconsistent with LCH. The point estimate is somewhat high, but the estimate is not very precise.

III. Comparison with Kormendi's Analysis

The results presented in column 1 of Table 1 are drastically different—in fact, just about diametrically opposite—to those reported by Kormendi in his various tests. This difference is clearly brought out by a comparison of the results of column 1 with those reported below, which rely on Kormendi's definition and measurement of variables (as nearly as we could match it) and on his precise specifications, including in particular the use of first differences. The only difference consists in the omitting of the years previous to 1952 (for which several of the series used in col. 1 are not readily available). The results of the replication are

$$
\text{(9)}\quad C^* = \underset{(0.01)}{0.02} + \underset{(0.07)}{0.27}Y_t^* + \underset{(0.04)}{0.02}Y_{t-1}^* + \underset{(0.01)}{0.01}A_t^*
$$

$$
- \underset{(0.05)}{0.09}G_t^* - \underset{(0.10)}{0.22}E_t^* + \underset{(0.12)}{0.16}T_t^* + \underset{(0.40)}{0.86}TR_t^*
$$

$$
R^2 = 0.72;\ SE = 0.02;\ D\text{-}W = 1.7;
$$

$$
\text{Log Likelihood} = 68.08,\ \text{Sample: } 1952\text{–}76.
$$

(The asterisks indicate that Kormendi's data definitions are used.)

It is readily apparent from a comparison of (9) with Kormendi's Table 5, that the results for the postwar period are substantially the same as those obtained by Kormendi for the longer period 1931–76 —the only significant exception being private wealth, which is substantially more important in his original period. In particular, with respect to the variables crucial for our

Table 2—Estimates of the Aggregate Consumption Function, 1952–76[a, b]

Independent Variables		(1)[c]	(2)[c]	(3)[c]	(4)	(5)
Constant	a	−0.394	−0.148	−0.238	−0.167	−0.121
		(0.152)	(0.222)	(0.114)	(0.127)	(0.141)
Net Worth (A)	b_0	0.031	0.030	0.031	0.030	0.032
Government Debt	b_1	0.073	0.106	0.096	0.096	
(G)		(0.058)	(0.063)	(0.042)	(0.046)	(0.051)
Net National	Σc_{1i}	0.896	0.835	0.876	0.882	0.763
Product (Y)		(0.028)	(0.055)	(0.021)	(0.076)	(0.046)
Net Taxes (T)	Σc_{2i}	−0.896	−0.835	−0.876	−0.840	−0.438
		(0.028)	(0.055)	(0.021)	(0.221)	(0.165)
Deficit (D)	Σd_i	0.016	−0.100	0.109	–	–
		(0.012)	(0.132)	(0.086)		
Government	Σd_i^*	–	–	–	−0.070	−0.062
Expenditure (E)					(0.154)	(0.132)
Transfers (TR)	Σt_i	–	0.013	–	–	–
			(0.118)			
Transitory	Σh_i	–	–	0.864	0.848	–
Taxes (T^*)				(0.200)	(0.367)	
SE		0.015	0.014	0.010	0.010	0.011
Log Likelihood		74.18	76.44	84.42	86.52	81.95
D-W		1.50	1.46	2.26	2.33	2.08

[a] See Table 1.

[b] The mean of the dependent variable is 2.89 and its standard deviation is 0.50.

[c] See Table 1.

tests, one finds that (*i*) the coefficient of taxes, instead of being strongly negative as in column 1, is positive (though insignificantly so); (*ii*) government expenditure, insignificant in column 1, is negative and rather significantly so; (*iii*) finally, since in (9) wealth (A^*) excludes government debt, the coefficient of that variable should be positive for LCH and zero for REP; it is actually negative and large—though it is rather imprecisely estimated. All of these results are, at least qualitatively, consistent with the "consolidated" REP hypothesis and inconsistent with LCH.

But a closer examination of (9) raises some serious questions as to the reliability of the results and their consistency with REP. To begin with, all the income and wealth variables have preposterously low coefficients—the long-run marginal propensity to consume with respect to income is less than .3(!), compared with an average propensity of over .9 and a customary estimate of the marginal propensity .7 to .9. Similarly, the wealth effect is but a fraction of the usual estimates and not even significant. Since none of the remaining fiscal variables are particularly significant, it is not surprising to find that the equation has quite a large standard error as compared with that of column 1 (or to others reported in Table 2 for the same period as (9)).

Looking more specifically at the fiscal variables, the first highly disquieting feature of (9) is that the coefficients completely fail to meet the consistency requirement that the sum of the coefficients of taxes and expenditure approximately match in magnitude that of income. Indeed, Kormendi's T and E coefficients add up to but −.06 compared with an already very low income coefficient of .29. Thus, according to (9), consumption weakly responds to income, and, in addition, responds even less to the loss of resources through a balanced-budget increase in taxes and government use of resources. Hence, an increased government use of resources would have to result in a large loss of net capital formation—clearly not what the REP model implies.

But what is most incongruous about equation (9) and other Kormendi results is the huge and even quite significant coefficient of transfer payments, TR^*. Clearly, transfer payments are part of net taxes, and by the REP hypothesis net taxes must have a near zero coefficient. Kormendi's explanation in terms of people hit by positive taxes not responding at all to the resulting loss of income (or even responding positively according to (9)), while people hit by negative taxes respond by a huge rise in consumption, is not only totally unconvincing but in addition, if true, would be as anti-REP a finding as one may conceive: it implies that those who receive a deficit-financed rise in transfer consume it merrily, thinking nothing of the effect on future generations, while all the rest equally fail to respond by cutting their consumption—with the result that the fall in net taxes would result in a rise in consumption, even though government expenditure were constant.

What accounts for the striking differences between the results of Kormendi and those reported in column 1? We can first investigate one obvious possibility, namely the difference in periods. Indeed, column 1 includes 8 years in addition to the 25 used in (9), and these years were characterized by extreme values of the deficit, not to mention variables like interest rates or inflation that could affect results. However, in Table 2, column 1 shows that estimating the basic equation (8) for the shorter period 1952–76, yields results very close to those of Table 1. None of the coefficients change significantly and, in particular, the coefficient of the deficit remains insignificantly different from zero.

We propose to show that the differences between (9) and the estimates of Table 1 are traceable in part to several errors in Kormendi's construction of the data and, to a much greater extent, to serious misspecifications of his equations.

There are three major flaws in Kormendi's data construction. First, he incorrectly deflates government expenditures, taxes, and transfers by a price index based on goods purchased by the government (in the case of government expenditures) or by the implicit price deflator for net national product. The correct price index, which gives the value of government spending, taxes, and transfers in terms of consumption goods, should be based on an appropriate consumption basket. Second, Kormendi creates his own measure of the service-flow measure of consumption, but does not use the analogous price index. (We use the service-flow measure of consumption developed by the Federal Reserve to support their quarterly econometric model.) Third, Kormendi uses an incorrect measure of real government interest payments. The correct measure should be based on the product of the real interest rate and the stock of debt; Kormendi instead deflates the reported nominal interest payments by the net national product price index. His procedure does not correct for change in real debt due to inflation (which corresponds also to the inflation premium built into the government's interest payments when Fisher's Law holds). The relevance of this correction for accounting purposes is pointed out by Stanley Fischer and Modigliani (1978); Modigliani, Tullio Jappelli, and Marco Pagano (1985) analyze its behavioral significance. (Our correction to the government's interest payments is based on the *ex post* interest rate.)

The role of these measurement errors can be assessed by comparing our (Kormendi's) equation (9) with the equation below which uses corrected data, but is identical to (9) in terms of specifications and sample period.

$$
\text{(9}'\text{)}\qquad \begin{aligned}
\Delta C = \underset{(0.01)}{0.02} + \underset{(0.08)}{0.24Y_t} + \underset{(0.05)}{0.11Y_{t-1}} + \underset{(0.01)}{0.01A_t} \\
+ \underset{(0.09)}{0.05G_t} - \underset{(0.14)}{0.04E_t} - \underset{(0.11)}{0.00T} + \underset{(0.27)}{0.77TR_t} \\
R^2 = 0.74;\ SE = 0.019; \\
D\text{-}W = 2.06;\ \text{Log Likelihood} = 66.2.
\end{aligned}
$$

(Note: *A* includes government debt.)

Equation (9′) shares with (9) the characteristic of little credibility—in particular, the coefficient of wealth and that of income remain unbelievably low, and the standard error is hardly smaller. However, with respect to the fiscal variables, even the limited support for REP that (9) appears to provide vanishes: the coefficient of debt becomes positive, though even less reliably estimated, and the coefficient of expenditure is essentially zero.

The remaining drastic difference between (9′) and Table 2, column 1, must clearly result from the unsatisfactory specification underlying Kormendi's tests. In the first place, once it is realized that the flow variables appearing on the right-hand side of (8) represent basically very long-run expectations, it becomes clear that it is entirely unsatisfactory to approximate these expectations just by the current value (or, at most, for income, current and lagged once); what is needed, at the very least, are reasonably long distributed lags. This conjecture is fully confirmed by the empirical evidence which also suggested five years as an adequate approximation.

A second shortcoming of Kormendi's specification is that he fails to impose the consistency condition requiring that the coefficients of taxes expenditure and income add up to zero. A final and more consequential shortcoming is that he chooses to estimate coefficients in the first difference form, a rather arbitrary procedure since, when he estimated an autoregression coefficient, its value, though fairly high (as Kormendi reports in his Table 1), is significantly below unity. But, in addition, the high autoregression coefficient he obtains is largely a predictable consequence of his failure to allow for sufficiently long distributed lags. With a five-year distributed lag, we find uniformly that *D-W* statistics are above the critical level; accordingly, all equations reported in Tables 1 and 2 have been estimated in level form by ordinary least squares (*OLS*).[4] These estimates were also checked using an instrumental variables procedure, but the results were not noticeably different from the *OLS* estimates and are therefore not reported separately.[5]

Replacing Kormendi's specification with ours not only reduces drastically the serial correlation of the errors, but, in addition, produces a substantial improvement in fit —the standard error of Kormendi's (9) is about 1/3 larger that those of Tables 1 and 2. Note that the improvement occurs even though Kormendi's equation includes, and ours excludes, the questionable variable transfer payments which is seen to contribute to his fit more than any other variable. But, as can be seen from columns 2 in Tables 1 and 2, our qualms about the role of this variable were justified: indeed, when transfer payments are added to the properly specified equations of both columns 1, the contribution of this variable, measured by the sum of the lagged coefficients, is essentially zero, implying that the effect of transfer payments on consumption is the same as that of any other component of net taxes. Finally, the coefficients of the traditional variables are reasonable and credible. We suggest, therefore, that Kormendi's results, and conclusions based thereon, can be dismissed without further ado.[6]

IV. Suggested Improvements in the Specification of Fiscal Variables

We would like to suggest some possible improvements in the specification of our equations of columns 1, Tables 1 and 2. That specification imposes the constraint that the sum of the coefficients of taxes expenditure and income be equal to zero at each of the five lags. This is unnecessarily restrictive; indeed, the shape of the distributed lags are intended to reflect the expectation formation mechanism, and there is no reason why this mechanism should be identical across different variables. This consideration is especially relevant with respect to taxes, since, as pointed out by Modigliani and Charles Steindel (1977), changes in the tax rate component of tax receipts are, typically, of a permanent nature. The above constraint can be readily eliminated by estimating separate distributed lags for income and each of the two fiscal variables. This procedure will, of course, also eliminate the valid "consistency" restriction that the overall sum of the coefficients, over all variables and lags, be close to zero. But, precisely for this reason, unconstrained estimation provides an opportunity to test whether the "consistency specification" is supported by the data.

---

[4] In every case where the *D-W* statistic approached the critical level we reestimated the equation using an autoregressive transformation; the autoregression coefficient was never significantly different from zero even at the 5 percent level, and the estimated coefficients were not appreciably different from those reported in the table.

[5] Lagged income and government expenditures were used as instruments for income and the deficit, while wealth and debt were treated as exogenous.

[6] Barro has pointed out (in private communication) that, in some cases, the coefficients on taxes in the consumption function may reflect their predictive power for future government spending. This would be the case if, as Barro has postulated (1979, 1984, 1985), the government attempts to smooth tax rates to reflect long-term government spending rather than transitory expenditures. If taxes are indeed better predictors of government spending then are past spending, the REP implies they should be negatively correlated with consumption. (Note that, as Modigliani, 1984b, points out, tax smoothing is not an implication unique to the REP.) This "tax smoothing" effect appears to play no significant role in our analysis. We tested the predictive power of lagged taxes for government expenditure by estimating a regression of government expenditure on goods and services on lagged (four years) expenditures and various measures of net taxes (taxes net of transfers, taxes net of transfers and government real interest payments, permanent net taxes and permanent taxes net of transfers and real interest). In no case did the tax measure contribute significantly to the explanatory power of the regression. The largest value of the test statistic for the restriction that taxes played no role in the forecasting equation was 4.63; the critical value at the 5 percent confidence level is 9.49. Lagged expenditure alone "explains" more than 98 percent of the variation in government expenditures on goods and services over the 1953–84 period.

A second shortcoming of the specification relates to the measurement of taxes. It is apparent from equation (6) that the income, tax, and expenditure terms in (8) are all meant to measure the average value (weighted for time and survival rates) of these variables expected by those currently living. This poses a problem when current and past values cannot be taken as representative of the future. A clear instance of nonrepresentativeness is the well-known case of tax changes which are announced, and presumably believed, to be but transitory. In the period covered by our sample there have been two important cases of transitory tax changes—one is the surtax of 1968–70, and the other the 1975 rebates. One can also conceive of significant transitory government expenditures (for example, major wars), but no important instance appears to have occurred during our sample period.

Since transitory tax changes should add relatively little to the present value of the weighted future tax liabilities, their effect on consumption should be much closer to zero than the permanent tax effect. To test this hypothesis, and hopefully reduce the unexplained error, one can continue to use the total tax variables of columns 1 and 2 in Table 1, but add the transitory tax component as a separate variable: the coefficient of this variable (h_i) will then measure the differential effect on consumption of transitory vs. permanent taxes.

The role of transitory taxes is tested for the entire period in columns 3 and 4 of Table 1, using the approach described above. In column 3, in order to focus on transitory taxes, we maintain the constraint on the sum of income, expenditures, and (permanent) tax coefficients. It is seen that the effect on consumption of temporary taxes is much less negative than that of permanent taxes—the estimated difference is 63¢ per dollar of taxes, and it is highly significant, resulting also in a striking improvement in fit and reduction in serial correlation of the residuals. This estimate implies a marginal effect of temporary taxes on consumption of only $-.48$ (with a standard error .2), and even this figure probably overestimates the negative effect, given that the point estimate of the effect of permanent taxes appears upward biassed (larger than -1). In column 4, we have, in addition, eliminated the "consistency" restriction, estimating separately coefficients for before tax income (Σc_{1i}), net taxes (Σc_{2i}), and government expenditure (Σd_i), as well as for transitory taxes (Σh_i).

It is apparent that the unrestricted coefficients of column 4 of Table 1 come very close to satisfying the consistency requirement, as the sum of the (permanent) tax and expenditure coefficients offset about exactly that of before tax income. Accordingly, as shown by a comparison of the likelihood statistic for columns 3 and 4, the restriction imposed in column 3 cannot be rejected at any reasonable level of significance. The only problem with the results of column 4 is that, in common with those of columns 1 through 3, they tend to give an unreasonably high point estimate of the effect of permanent taxes, matched by an equally questionable positive effect of government expenditure on consumption. However, this may be accounted for by imprecision of the estimates, as measured by their standard error. In particular, the tax coefficient differs from the income coefficient by but one standard error and the same is true of the difference of the expenditure coefficient from zero.

For the sake of completeness, the test of columns 3 and 4 of Table 1 are reported for the shorter period in Table 2. The estimate of the temporary tax effect for this period must be regarded as less reliable because only a portion of the experience of the 1975 tax rebate is reflected in the data. Nonetheless, the estimates are seen to confirm that the effect of transitory taxes on consumption is a great deal smaller than that of permanent taxes. Furthermore, the difference (even greater than in cols. 3 and 4) is highly significant, resulting, again, in a considerable improvement in fit. In Table 2, column 4 confirms again that the constraint imposed on column 3 is not rejected by the data—indeed, the coefficients of column 3 are all very similar to those of column 2. It is also encouraging that the permanent tax coefficient is of very reasonable magnitude both absolutely and relative to that of income, and that the coefficient of expenditure has the expected negative sign, though small, as implied by LCH.

Column 5 of Table 2 reports one more result which we find instructive. The coefficients in this column are estimated without constraint but also omitting the temporary taxes. It appears that failing to allow for the much smaller impact of transitory taxes, seriously biases up (i.e., towards zero) the coefficient of the tax variable. As a result, the sum of the coefficients of taxes and expenditure is but $-.5$, compared with an income coefficient of .76. It is not unlikely that failure to allow for transitory taxes may be one of the reasons why Kormendi could not find a negative effect of taxes on consumption and obtained coefficients which completely fail to meet the consistency test.

There is one more empirical result that we feel compelled to report. It was shown in Tables 1 and 2, columns 2, that transfer payments, when added as an additional variable, receive a coefficient near zero. Yet, when transfer payments are added to column 3 of Table 1—which separates transitory from permanent taxes—the effect is truly devastating: the coefficient of transfers is .54 with t-ratio of 5(!); since this variable is already included in income minus net taxes (coefficient .77) and the deficit (coefficient .23), the total implied effect of transfers comes to 1.54. For the short period, the results are about as incredible, as the total effect of transfers is 1.36 with a t-ratio of 2.5. In addition, the estimated differential effect on consumption of transitory vs. permanent taxes rises in Table 1, column 3, from .63 to 1.4; since the effect of permanent taxes implicit in the coefficient of income and deficit is 1.0 (already rather unreasonably large), the effect on consumption of temporary taxes would be 0.4; that is, a dollar of temporary taxes would *increase* consumption by 40¢. One further surprise is that the residuals exhibit a marked negative serial correlation, a rather suspicious finding for these kind of time-series regressions.

One explanation for the significance of transfers in the consumption function is that they may provide a reliable signal of one component of permanent income. In particular, Social Security and other transfers to the elderly may be more certain than labor income, which may be reduced or eliminated by unexpected retirement. That informational role may be enhanced in economic downturns, which could account for the peculiar interaction with temporary taxes. Nevertheless, we found no substantial improvement in the results when transfers were replaced by social security payments.

On the whole, these results appear so preposterous that, in our view, they should be disregarded—although they also appear so unlikely to reflect pure chance that they deserve to be filed under agenda for the future.

V. Conclusions

We have derived an aggregate consumption function based on the life cycle paradigm and have shown that it implies that consumption should be affected both by taxes and by government expenditure and by private claims on the government as well as the wealth. It was shown that the relative significance of taxes and government expenditure in the consumption function reflects mortality rates, the length of life, and both the personal discount rate and interest

TABLE 3—DATA USED IN THE REGRESSIONS[a]

Year	C	A	G	Y	TR	T	E	D	T^*
1949	2.09830	10.24789	2.54671	2.94840	0.14645	0.48932	0.47835	−0.01097	0.00000
1950	2.16033	10.31647	2.56432	3.13978	0.17216	0.72363	0.45953	−0.26410	0.00000
1951	2.20120	10.45613	2.26725	3.31530	0.12779	0.89179	0.65505	−0.23674	0.00000
1952	2.24048	10.88833	2.12983	3.31906	0.12572	0.82212	0.78714	−0.03498	0.00000
1953	2.27847	10.94125	2.12317	3.37896	0.13028	0.81541	0.83085	0.01544	0.00000
1954	2.29297	10.94416	2.15840	3.25165	0.14768	0.70030	0.73847	0.03817	0.00000
1955	2.36421	11.54037	2.15594	3.44293	0.15272	0.79534	0.70773	−0.08761	0.00000
1956	2.42094	12.16621	2.08327	3.47656	0.15750	0.84806	0.72193	−0.12613	0.00000
1957	2.44090	12.25389	1.98322	3.48123	0.17389	0.83292	0.75576	−0.07716	0.00000
1958	2.45015	12.09346	1.95084	3.39446	0.20396	0.73773	0.79536	0.05764	0.00000
1959	2.50200	12.98754	2.00864	3.56677	0.20311	0.84357	0.78826	−0.05531	0.00000
1960	2.53064	13.33399	1.96582	3.56636	0.20892	0.87105	0.77912	−0.09193	0.00000
1961	2.55696	13.22858	1.93199	3.57854	0.23107	0.82274	0.81353	−0.00921	0.00000
1962	2.60643	13.93548	1.95742	3.76567	0.23059	0.89504	0.86360	−0.03144	0.00000
1963	2.65242	13.59904	1.94225	3.86097	0.23651	0.93752	0.87738	−0.06014	0.00000
1964	2.74833	14.24857	1.90374	4.02640	0.23926	0.92876	0.89661	−0.03215	0.00000
1965	2.84715	15.05566	1.89709	4.23856	0.25113	1.00115	0.92685	−0.07431	0.00000
1966	2.95633	15.37331	1.85059	4.43978	0.26568	1.10352	1.01637	−0.08716	0.00000
1967	3.03115	14.80681	1.81809	4.51820	0.30450	1.09996	1.11343	0.01347	0.00000
1968	3.14352	15.45900	1.84912	4.70265	0.33124	1.23030	1.17176	−0.05854	0.05166
1969	3.24328	16.28678	1.81697	4.80729	0.34923	1.33364	1.16609	−0.16755	0.06845
1970	3.31416	15.64229	1.74252	4.75079	0.39903	1.21346	1.16140	−0.05206	0.04359
1971	3.36932	15.80857	1.79200	4.90640	0.45008	1.17574	1.18371	0.00797	0.00000
1972	3.49398	16.42354	1.86056	5.14208	0.47539	1.29634	1.21862	−0.07772	0.00000
1973	3.59773	17.64463	1.82951	5.38611	0.50738	1.41186	1.22128	−0.19058	0.00000
1974	3.61446	17.10263	1.67637	5.21342	0.54295	1.40182	1.23742	−0.16441	0.00000
1975	3.67342	16.26371	1.64065	5.16574	0.63499	1.15164	1.28036	0.12872	−0.06880
1976	3.80017	17.63173	1.86191	5.44617	0.65774	1.24486	1.29434	0.04948	0.00000
1977	3.92910	18.68652	1.94918	5.72281	0.66191	1.38205	1.32615	−0.05590	0.00000
1978	4.05878	18.82615	1.97971	5.93879	9.65673	1.52707	1.34803	−0.17904	0.00000
1979	4.15649	19.90689	1.85136	5.97227	0.66265	1.55996	1.34092	−0.21904	0.00000
1980	4.19528	19.94764	1.72724	5.77520	0.70522	1.45075	1.35888	−0.09187	0.00000
1981	4.23351	20.73876	1.71721	5.87696	0.72533	1.40111	1.37129	−0.02981	0.00000
1982	4.26121	20.65688	1.77052	5.66342	0.75579	1.21792	1.44791	0.23000	0.00000
1983	4.36233	21.23510	2.06152	5.83235	0.77590	1.22102	1.45229	0.23127	0.00000
1984	4.49736	21.91946	2.29538	6.15170	0.75358	1.31457	1.44690	0.13233	0.00000

[a] All data are in real thousands of 1972 dollars per capita terms.

rate, but that as long as the representative planning horizon does not extend significantly beyond life, the role of taxes could be expected to be substantially larger than that of expenditure and the role of debt close to that of other wealth.

We have endeavored to show that Kormendi's contrary empirical findings—that consumption does not respond to taxes but only to government expenditure—are the result of errors in the measurement of variables and, to a greater extent, the consequence of serious misspecifications of his test equations. Specifically, Kormendi fails to allow for the relatively long lags required to approximate permanent income, or to distinguish between permanent and temporary tax effects and relies on an inefficient method of estimation.

We have provided evidence that, once Kormendi's measurement errors and misspecifications are corrected, the data for the United States in the postwar period are strikingly and unmistakenly consistent with a "Life Horizon"-Life Cycle approach to consumption behavior and equally inconsistent with the infinite horizon-REP formulation.

DATA APPENDIX

The data used in this paper are from the Federal Reserve data bank. A detailed specification of the data is available in documentation accompanying the Federal Reserve data tapes. All data are per capita and expressed in thousands of 1972 dollars.

Consumption: The sum of expenditures on nondurable goods and services plus a fraction of expenditures or durables and a measure of the flow of services from the stock of consumer durables.
Price Index: A consumer price index based on prices of nondurables, services and durable goods weighted according to their proportions of consumption.
Income: Net national product.
Taxes: Total federal plus state and local taxes (including employer's contribution to social security), net of transfers and real interest payments. Real interest payments are calculated by subtracting the product of the rate of inflation and the beginning-of-period government debt from actual total government (including the Federal Reserve) net interest payments. Our estimate of temporary taxes is taken from the Economic Reports to the President of 1969 and 1975.
Government Expenditures: Federal and state and local expenditures on goods and services.
Wealth: Net worth of households.
Debt: The sum of federal and state and local net financial liabilities.

The data used in the regressions are reported in Table 3.

REFERENCES

Aschauer, David Alan, "Fiscal Policy and Aggregate Demand," *American Economic Review*, March 1985, *75*, 117–27.

Barro, Robert J., "Are Government Bonds Net Wealth?," *Journal of Political Economy*, November/December 1974, *82*, 1094–117.

______, "On the Determination of the Public Debt," *Journal of Political Economy*, October 1979, *87*, 940–71.

______, "U.S. Deficits Since World War I," paper presented at Conference on Growth and Distribution: Intergenerational Problems, Uppsala Sweden, June 1984; *Proceedings* forthcoming.

______, "Government Spending, Interest Rates, Prices and Budget Deficits in the United Kingdom, 1730–1918," Working Paper No 1, University of Rochester, March 1985.

Barsky, Robert, "Three Interest Rate Paradoxes," unpublished doctoral dissertation, MIT, 1985.

Blanchard, Olivier J., "Debt, Deficits and Finite Horizons," *Journal of Political Economy*, April 1985, *93*, 223–47.

Feldstein, Martin S., "Government Deficits and Aggregate Demand," *Journal of Monetary Economics*, January 1982, *9*, 1–20.

Fischer, Stanley and Modigliani, Franco, "Towards an Understanding of the Real Effects and Costs of Inflation" *Weltwirtschafliches archiv* (*Review of World Economics*), No. 4, 1978, *114*, 736–81.

Hayashi, Fumio, "The Permanent Income Hypothesis: Estimation and Testing by Instrumental Variables," *Journal of Political Economy*, October 1982, *90*, 895–916.

Kormendi, Roger C., "Government Debt, Government Spending and Private Sector Behavior," *American Economic Review*, December 1983, *73*, 994–1011.

Modigliani, Franco, (1984a) "The Economics of Public Deficits," in *Economic Policy in Theory and Practice* (*Proceedings* of 1984 Pinhas Sapir Conference on Development in Memory of Abba P. Lerner, May 28–30, 1984, Tel-Aviv University), New York: McMillan, forthcoming.

______, (1984b), "Comment on 'U.S. Deficits Since World War I' by R. J. Barro," presented at Conference on Growth and Distribution: Intergenerational Problems, Uppsala Sweden, June 1984; *Proceedings* forthcoming.

______ **and Steindel, Charles,** "Is a Tax Rebate an Effective Tool for Stabilization Policy?," *Brookings Papers on Economic Activity*, 1:1977, 175–209.

______, **Jappelli, Tullio and Pagano, Marco,** "The Impact of Fiscal Policy and Inflation on National Saving: The Italian Case," *Banca Nationale del Lavoro Quarterly Review*, June 1985, No. 153, 91–126.

Seater, John J. and Mariano, Roberto S., "New Tests of the Life Cycle and Tax Discounting Hypotheses," *Journal of Monetary Economics*, March 1985, *15*, 195–215.

Sterling, Arlie, "Public Debt and Private Wealth: The Role of Anticipated Taxes," unpublished doctoral dissertation, MIT Sloan School of Management, April 1985.

Tobin, James, *Asset Accumulation and Economic Activity* (Yrjo Jahnssen Lectures), Chicago: University of Chicago Press, 1980.

Government Debt, Government Spending, and Private Sector Behavior: Reply

By ROGER C. KORMENDI AND PHILIP MEGUIRE*

The effects of fiscal policy on private sector behavior has long been an important and controversial issue in macroeconomics. That this is still the case is attested to by the length and content of the comments by James Barth, George Iden, and Frank Russek (BIR), and by Franco Modigliani and Arlie Sterling. We address these comments in turn.

I. Barth, Iden, and Russek

Barth et al. make two significant contributions. First, they undertake a replication of Kormendi's (1983) variables and estimates. We especially admire their meticulous reconstruction from original sources of Kormendi's data set. Although they find some minor errors in his data, they are able to replicate his results quite closely (Table 1, BIR) and show that his results continue to hold when the years 1977–83 are added to the sample (Table 2, BIR). Such painstaking replication, too infrequent in economics, should be encouraged in any scientific discipline.

Second, Barth et al. explore whether Kormendi's results favoring what he termed the "Consolidated Approach" to the analysis of the government and private sectors, are robust to 1) disaggregating government debt into its federal and state and local components; 2) valuing government debt at par instead of market; and 3) varying the sample period. Taken one at a time, these explorations support the Consolidated Approach. For certain postwar subperiods, however, Barth et al. do find that specifications incorporating either 1) or 2), occasionally yield coefficients not predicted by the Consolidated Approach. They do not claim too much for these problematic results, but suggest that their results do indicate directions for further research.

We propose to interpret BIR's results somewhat differently. In particular, when testing whether a given specification is sensitive to alternative choices of explanatory variables or sample periods, one should *nest* the specifications derived under both the null (benchmark) and the alternative hypotheses into a general specification and then test whether the null is rejected. While Barth et al. do not generally undertake such tests, they do report the summary statistics needed to compute them. Table 1 presents the summary and test statistics implied by a pertinent subset of BIR's regressions. In all cases, the benchmark regression is (7) from BIR's Table 2 in Kormendi's "augmented" Consolidated Approach specification estimated over the entire 1931–83 period.[1] For this benchmark and each of the five alternatives, we report the standard error of the regression (*SER*), the degrees of freedom (*DOF*), and the implied sum of squared residuals (*SSR*), calculated as $DOF \times SER^2$. The columns 8 and 9 present the F (Chow) statistics, along with their marginal significance levels (*MSL*s), for testing the benchmark against the alternatives. In all cases, the tests fail to reject the consolidated benchmark at standard significance levels.

The only subperiod analyzed in Table 1 is 1947–83 (or the closest reported alternative), while Barth et al. report estimates drawn from a number of other postwar subperiods. For any given significance level, the probability of falsely rejecting the null (the benchmark applied to the entire sample) increases with the number of subperiods explored. Hence formal inferences can be drawn only from regressions estimated over

---

*University of Michigan, Ann Arbor, MI 48109-1234.

[1] Using (6) from this table instead of (7) would not materially alter the conclusions that follow.

TABLE 1—*F*-TESTS OF BENCHMARK CONSOLIDATED SPECIFICATION AGAINST ALTERNATIVES SUGGESTED BY BARTH ET AL.

Equation in BIR (1)	Sample Period (2)	Federal/ State and Local? (3)	Par-Market? (4)	*SER* (5)	*DOF* (6)	*SSR* (7)	*F* (Chow) Statistic (8)	Numerator Degrees of Freedom (9)	Marginal Significance Level (10)
Benchmark									
7	1931–83	No	No	0.0213	43	0.0195	–	–	–
Alternatives									
17	1947–83	No	No	0.0207	27	0.0116	1.1	16	0.40
10	1931–83	Yes	No	0.0215	42	0.0194	0.2	1	0.66
19	1947–83	Yes	No	0.0185	26	0.0089	1.8	17	0.09
43	1931–83	No	Yes	0.0212	42	0.0189	1.3	1	0.26
45	1949–83	No	Yes	0.0194	24	0.0090	1.5	19	0.17

Notes: Col. 5: *SER* is the standard error of the regression identified in col. 1; col. 6: *DOF* is the degrees of freedom of this regression (also the denominator degrees of freedom for the *F* (Chow) statistic); col. 7: $SSR = DOF \times SSR^2$. A "Yes" in col. 3 means that total government debt has been disaggregated into its Federal and State and Local components; a "Yes" in col. 4 means that a variable measuring the difference between market and par valuation of total government debt has been added to the regression.

subperiods that are well-motivated a priori, such as 1947–83 which separates World War II and the Great Depression from the remainder of the sample. This has guided our choice of regressions from among the many reported by Barth et al. In view of the *MSL*s reported in Table 1, we find BIR's results to be supportive of the Consolidated Approach.[2]

II. Modigliani and Sterling

Modigliani and Sterling "rely on the Life Cycle Hypothesis" (LCH) to derive an empirical consumption relation which they claim admits the Ricardian Equivalence Proposition (REP) as a limiting case. We show that their LCH specification embodies implicit untested restrictions that prevent them from nesting the competing hypotheses of interest. Releasing these LCH restrictions produces a nested specification similar to the one set out in Kormendi (1983). Estimating this nested specification over both Modigliani and Sterling's data and our own rejects the restrictions implied by their formulation. Modigliani and Sterling's use of distributed lags and undifferenced data to estimate their empirical model is also questionable in view of the nonstationarity typical of macroeconomic time-series data. To address this point, we extend the sample back to 1929 and find that both the restrictions and the coefficient homogeneity of Modigliani and Sterling's LCH specification are rejected, whereas those of the consolidated specification are not.

A. *A General Specification for Nesting the LCH and Consolidated Approaches*

One of the main contributions of Kormendi (1983) was to show how to test what he termed the standard and consolidated approaches as special cases of a general specification. Modigliani and Sterling's LCH and the Consolidated Approach can be nested in a similar manner. First, express the government deficit, D, in terms of the government budget constraint to obtain[3]

$$(1) \qquad D = E + TR + RGINT - TX.$$

---

[2] Even regressions with problematic coefficients for federal vs. state and local debt or par vs. market value of debt typically have coefficients for taxes and government spending that support the Consolidated Approach.

[3] Following the notation of Modigliani and Sterling where possible, E denotes government spending, TX *gross* taxes, TR transfer payments, and $RGINT$ *real* interest payments on government debt, i.e., the interest on government debt net of any inflationary capital gains or losses.

Substituting (1) and the definition of Net Taxes $T = TX - TR - RGINT$ into Modigliani and Sterling's equation (8) yields[4]

$$\begin{aligned}(2)\quad C_t = a &+ b_0 A_t + b_1 G_t + c_1(L) Y_t \\ &+ c_2(L) E_t + c_3(L) TX_t \\ &+ c_4(L) RGINT_t + c_5(L) TR_t + e_t\end{aligned}$$

It is important to note that two sets of restrictions have now been released—those implied by what Modigliani-Sterling term the "Consistency Conditions," and those implied by their concept of Net Taxes. By releasing these restrictions and allowing each lag polynomial $c_i(L)$ to be independently estimated, we derive the nested specification for testing the restrictions implicit in their formulation. Under their LCH, $c_2(L) = 0$, and the Consistency Conditions require that $c_1(L) = -c_3(L) = c_4(L) = c_5(L) > 0$. Under the Consolidated Approach, $c_3(L) = c_4(L) = 0$ is the key null hypothesis, and $c_2(L) < 0$ would also hold if government spending is either a substitute for private consumption or is dissipative.[5]

Difficulties with Modigliani and Sterling's restrictions arise from two sources. The first concerns their transition from their theoretical equation (6), which is in terms of *expected future* values, to their empirical specification (8), which is in terms of a distributed lag on *realized past* values. From Thomas Sargent (1979) and others, we know that distributed lag coefficients will reflect the time-series properties of the explanatory variables (through the expectations formation process), as well as the coefficients of the expected future values.[6] Therefore, Consistency Conditions appropriate for a model stated in terms of expected future values will not generally apply to the coefficients of the analogous distributed lag model. Second, Modigliani and Sterling's formulation does not allow government spending to substitute directly for either private consumption or investment, nor does it allow transfers to have redistributive consequences for private consumption. If their theoretical development were extended to incorporate these effects, the Consistency Conditions would not hold even for a model stated in terms of expected future values.

The main difference between (2) above and (6) in Kormendi (1983) is that the former is a (polynomial) distributed lag in the levels of the (flow) variables, while the latter differences all variables and contains (almost) no lags. The issue of specification in levels vs. differences is of fundamental importance in analyzing macroeconomic time-series. C. W. J. Granger and Paul Newbold's (1974) classic paper shows that failure to difference nonstationary time-series data can produce spuriously large and significant regression coefficients.[7] Charles Plosser and G. William Schwert (1977), on the other hand, show that overdifferencing mainly reduces estimation efficiency. Finally, Robert Hall (1978) shows that on theoretical grounds one should expect consumption to be nonstationary. For these reasons, and because his empirical diagnostics showed strong evidence of nonstationarity, Kormendi chose first differences as his main specification. Kormendi also discussed the appropriate role of lags as reflecting expected future values of explanatory variables, implementing them in his equation (5) and in the results reported in his Table 3.[8] He could not reject, however, the hypothesis that the coefficiencts of lagged variables were zero.

---

[4] Following Modigliani-Sterling, C denotes consumption, A private wealth, G government debt. Y income, and $c_i(L) = c_{i0} + c_{i1}L + \cdots + c_{i4}L^4$.

[5] If government spending consisted mostly of investment, then $c_2(L) = 0$ could obtain under the Consolidated Approach as well. Hence the Consolidated Approach in no way requires that $c_1(L) = -c_2(L)$ as Modigliani and Sterling allege.

[6] David Aschauer (1985), in the context of a derivation similar to Modigliani-Sterling, contains a clear discussion of these points, along with empirical results incorporating them.

[7] N. Gregory Mankiw and Matthew Shapiro (1985) show that consumption will reveal a spurious "excess sensitivity" to income if consumption and income are in fact nonstationary and the consumption relation is inappropriately estimated over levels data. See also Kormendi and Laura LaHaye (1986) for a related discussion and empirical evidence drawn from 30 countries.

[8] His formulation was based on Hall and Sargent (1978), and is also related to Aschauer.

B. *Nested Tests of the LCH vs. the Consolidated Approach*

In order to focus on their implicit LCH restrictions, we temporarily set aside the problem of nonstationarity and grant Modigliani-Sterling their specification of polynomial distributed lags over levels (PDL/levels), as well as their variable definitions.[9] Table 2 contains an estimate of (2) using exactly Modigliani-Sterling's data, and should be compared with their Table 1, column 1.[10] Note first that the lag sums for the components of Net Taxes (*TX*, *TR*, and *RGINT*) differ radically from each other and from that of income (*Y*). In particular, the lag sum for *TX*, while negative, is small and not significantly different from 0. Moreover, all of the negative effect of taxes on consumption is concentrated in the (insignificant) third and fourth lags, so that the lag sum through two lags for *TX* is .06. Government spending, *GS*, on the other hand, yields a significant lag sum of −.27 through two lags, and is marginally significant through four lags only because the coefficients of the third and fourth lags are near zero. Finally, an *F*-test of the restrictions Modigliani-Sterling implicitly imposed on (2) rejects these at better than the .001 level.

TABLE 2—TEST OF THE IMPLICIT RESTRICTIONS IN MODIGLIANI AND STERLING'S LCH SPECIFICATION

Sample Period	1952–84
Polynomial Lag on Flow Variables:	
Lag Length	4
Degree	2
Constant	−0.35 (0.10)
A	0.010 (0.006)
G	0.13 (0.03)
Y	0.78 (0.08)
TX	−0.12[a] (0.19)
TR	0.65 (0.13)
RGINT	1.85 (0.28)
GS	−0.32[b] (0.20)
SER	0.010
D-W	2.5
Q(6)	6.4
F-Statistic for LCH Restrictions	5.9
DOF	(6,20)
MSL	0.001

Notes: All data, including the dependent variable *CON*, are from Modigliani and Sterling. *SER* is the standard error of the regression, *D-W* is the Durbin-Watson statistic, *Q*(6) is the Box-Pierce statistic computed from the first 6 residual autocorrelations. *MSL* is the marginal significance level of the *F*-statistic under the null distribution with (numerator, denominator) degrees of freedom (*DOF*). Standard errors of the estimated coefficients are shown in parentheses.

[a] Lag sum (standard error) through two lags = 0.06 (0.11).

[b] Lag sum (standard error) through two lags = −0.27 (0.15).

---

[9] Modigliani-Sterling claim to uncover three "serious flaws" in Kormendi's (1983) specification and data definitions: 1) an improper measure of consumption flow; 2) an improper method of deflating the regressors: and 3) an inappropriate use of nominal instead of real government interest payments. The Modigliani-Sterling consumption flow measure uses virtually the same formula as Kormendi but applies it to the FRB measure of the stock of consumer durables, whereas Kormendi used the BEA measure which is derived using considerably more detailed procedures. The choice of deflator is a typical "index number problem" and cannot be resolved a priori. We accept the real interest rate variable as the correct measure, but note that it should be computed using the *market* value of government debt (see also Kormendi, 1986). We point out here that, unlike Barth et al., the Modigliani-Sterling replication of Kormendi's results involves neither a specification nor a time period to be found in Kormendi (1983), nor are their variables sufficiently documented to trace their construction. As it turns out, using Modigliani-Sterling's dependent variable and deflation method does not materially affect the Barth et al. replication, and the use of real government interest payments actually improves the results from the perspective of the Consolidated Approach. See Table 4, col. 2, and fn. 15.

[10] Following Modigliani-Sterling, we only report the sums of the coefficients of the contemporaneous and lagged values (the lag sums), along with the standard errors of these sums. Although we refer only to col. 1, Table 1, in Modigliani-Sterling, similar results obtain for all their reported equations.

C. *Nonstationarity and Homogeneity over the 1931–83 Sample Period*

A regression model estimated over all of the available data that fails to reject a Chow test for equality of coefficients when the sample is divided, satisfies what we call a *homogeneity condition*. Such a test can also be interpreted as a general test for specification error.[11] In particular, if one estimates a regression over data in levels when nonstationarity mandates estimation over differenced data, then one would expect the homogeneity condition not to hold, especially for longer samples.

One of the main features of the differenced consolidated specification in Kormendi (1983) was that it satisfied this homogeneity condition when the full sample was divided into a "turbulent" war and Great Depression period and a "tranquil" postwar period. The Chow test reported in the second row of Table 1 above confirms that this homogeneity condition also holds for the 1931–83 period. We will now test whether Modigliani and Sterling's PDL/levels specification satisfies this homogeneity condition, splitting the sample in 1952, the first year in their sample. The chief obstacle to implementing such a test is that strictly comparable data for some of the Modigliani-Sterling variables are not available for years prior to 1948. We surmount this obstacle by applying their variable definitions and concepts as closely as possible to Barth et al.'s and Kormendi's data, which go back to 1929.[12]

TABLE 3—HOMOGENEITY TESTS FOR LCH SPECIFICATION, SHORT VS. LONG SAMPLE PERIOD

Sample Period	1952–83	1933–83
Constant	−0.21	0.32
	(0.10)	(0.02)
W	0.019	0.037
	(0.004)	(0.004)
EQ	0.033	0.045
	(0.007)	(0.011)
GB	0.14	0.026
	(0.03)	(0.018)
Y^a	0.79	0.52
	(0.05)	(0.05)
T^a	−0.79	−0.52
	(0.05)	(0.05)
D	0.29	−0.13
	(0.12)	(0.04)
SER	0.017	0.032
D-W	1.3	0.9
Q(6)	18.8	17.6
F	–	6.8
DOF		(19,24)
MSL		< 0.001

Notes: All data except for *EQ* are described in Barth et al. All variables are real per capita levels. *EQ* is the market value at year end of stocks listed on the New York Stock Exchange; *W*, *GB* and the dependent variable *PC* are defined in Barth et al.; *Y*, *T* and *D* are defined in Modigliani-Sterling. The coefficients of *Y*, *T*, and *D* are sums of the contemporaneous and 4 lagged coefficients, estimated from first-degree PDLs. The *F* statistic in the second column tests whether the coefficients in the two regressions are equal. For remaining abbreviations, see Notes to Table 2. Standard errors are shown in parentheses.

[a]The coefficients of *Y* and *T* have been constrained to be equal with opposite sign.

Column 1 in Table 3 shows that regression (1) in Table 1 of Modigliani-Sterling can indeed be well replicated using variables constructed from our data set. Column 2 of this table extends the sample back to 1931. The coefficient of government debt (*GB*) becomes insignificant and small. The lag sum of deficits (D) becomes significantly negative, which is consistent with the REP. The lag sums for income (Y) and net taxes (T) drop considerably. The very low Durbin-Watson statistic suggests that the residuals may even be nonstationary, in which case estimation in levels would be a serious specification error. Finally, the F-statistic and corresponding MSL at the bottom of Table 3 decisively reject homogeneity.[13]

---

[11] Such a test can be viewed as a member of the class of tests for specification error discussed by Jerry Hausman (1978).

[12] The two sets of variables are related as follows. The two consumption variables differ in their imputed flow of services from the stock of consumer durables. Kormendi's formula used the BEA's estimate of the net stock of consumer durables, while Modigliani-Sterling applied virtually the same formula to the less-detailed FRB estimate of this stock. Kormendi's variable *W* is an estimate of real net private tangible wealth valued at replacement cost plus human wealth measured at the cost of production. The Modigliani-Sterling variable *A*, on the other hand, measures the net worth of private households. Since the latter includes corporate equity, we have added the variable *EQ*, the market value of all stocks listed on the New York Stock Exchange, to our regressions. For the overlapping years, the data for real income and net taxes in the two data sets are virtually identical. We have followed Modigliani-Sterling in deflating all regressors by the same deflator as the dependent variable. Our results are estimated through 1983 instead of 1984 because we use Barth et al.'s extension of Kormendi's (1983) data set. Our data and variable definitions are further described in an Appendix, available upon request, which should be read jointly with the Data Appendix to Barth et al. (1986).

The results in Table 4 allow a direct comparison of PDL/levels *without* the Consistency Conditions (col. 1), with Kormendi's "augmented" consolidated specification estimated with differences/no lags (col. 2).[14] Note that with the Consistency Conditions removed, the lag sum of income (Y) is in the range found by Kormendi and Barth et al., that of taxes (TX) is insignificantly positive, that of government spending (GS) is negative (and significant after two lags), and the coefficient of government bonds (GB) is insignificant. Only the lag sum of real interest payments ($RGINT$) is at odds with the Consolidated Approach. However, specification problems remain evident. Both the D-W statistic and the $Q(6)$ statistic indicate that nonstationary residuals may still be a problem, while the Chow test again decisively rejects homogeneity.

TABLE 4—HOMOGENEITY TEST OF UNRESTRICTED PDL/LEVELS VS. AUGMENTED CONSOLIDATED SPECIFICATION

Sample Period	1933–83	1931–83
Lags	PDL	Y_{t-1} only
Differences:	No	Yes
Constant	0.66	0.010
	(0.11)	(0.005)
W	0.017	0.016
	(0.011)	(0.006)
EQ	0.029	–
	(0.013)	
GB	0.028	−0.019
	(0.030)	(0.020)
Y_t	0.36	0.30
	(0.11)	(0.05)
Y_{t-1}	–	0.07
		(0.03)
TX	0.13	0.02
	(0.26)	(0.09)
GS	−0.11	−0.23
	(0.06)	(0.03)
TR	1.04	0.82
	(0.44)	(0.19)
$RGINT$	0.83	0.07
	(0.23)	(0.08)
RE	–	0.08
		(0.14)
SER	0.025	0.022
D-W	0.8	1.3
$Q(6)$	27.7	5.4
F	4.3	1.1
DOF	(19,18)	(21,22)
MSL	0.002	0.41

Notes: All data and variables, including the dependent variable PC, are as described in Barth et al., or in the Notes to Table 3. In the first column, the reported coefficients (other than for W, GB, and EQ) are sums of the contemporaneous and 4 lagged coefficients, estimated from first-degree PDLs. The F-statistic tests whether the homogeneity condition holds when the sample is split in 1952. For remaining abbreviations, see Notes to Table 2. Standard errors are shown in parentheses.

Comparing Kormendi's differenced consolidated specification in column 2 with column 1 reveals some interesting differences. The $Q(6)$ statistic is now consistent with white noise residuals. The coefficient of government debt (GB) becomes negative. The coefficient of government spending (GS) becomes more negative and overwhelmingly significant. Of particular interest is the coefficient of *real* government interest payments ($RGINT$) which is now effectively zero. Thus the key null hypothesis of the Consolidated Approach, namely that the coefficients of TX, $RGINT$ and RE should all be zero, finds strong support in the data. Finally, the F-statistic at the bottom of column 2 supports the homogeneity of the consolidated specification.[15]

---

[13] Correcting for the serial correlation in the residuals by using the AR1 maximum likelihood procedure in TSP does not alter this F test or the associated coefficients materially.

[14] The data and specification are the same as those used to estimate Barth et al.'s equation (7), with the following exceptions where we follow Modigliani-Sterling instead. Nominal government interest payments ($GINT$) have been replaced by real ones ($RGINT$). The Federal Reserve System has been shifted from the private to the government sector, affecting both taxes (TX) and real interest payments ($RGINT$).

III. Conclusion

To summarize, we admire Barth, Iden, and Russek's meticulous replication and sample extension of Kormendi (1983). However, we show that Kormendi's benchmark consolidated specification estimated over the 1931–83 period is not rejected when pitted against those combinations of explanatory variables and sample periods with clear a priori motivation. With respect to Modigliani and Sterling, we show that their formulation imposes implicit restrictions that are rejected by the data. Moreover, in contrast to the Consolidated Approach, their conclusions do not hold when their sample is extended to include the Great Depression and World War II years.

We conclude by showing how one might synthesize the Consolidated Approach, the Life Cycle Hypothesis (LCH), and the empirical results presented as part of this debate. In this regard, consider a world with a given age distribution of agents, all of whom rationally forecast the future consequences of current debt-financed tax reductions, as per the REP. Suppose, moreover, that these agents have limited bequest motives, as per the LCH.[16] Then, if the median remaining life expectancy is sufficiently long, a general debt-financed tax cut falling over the entire age distribution would be virtually capitalized in full, and the coefficients of taxes and government spending would conform to the Consolidated Approach. However, to the extent that transfer payments redistribute wealth from "young" to "old" (for example, Social Security) life cycle considerations would imply a positive coefficient for transfers.[17] We find the policy implications of the foregoing to be potentially quite interesting.[18]

---

[15] Some additional observations are in order. First, the results favoring differences/no lags are not the result of a failure to include lags of the explanatory variables. As in Kormendi (1983), the data here also fail to reject the null hypothesis that all lags have zero coefficients ($F(13,29)=1.1$, $MSL=.40$). Second, *contra* the allegation of Modigliani-Sterling, deflating all explanatory variables by the same deflator as the dependent variables has no material effect on the estimates. Third, adding the variable *EQ* to this regression yields a marginally significant coefficient with no material effects on the other results. Fourth, Modigliani and Sterling argue forcefully that a value of .37 for the sum of the coefficients of income and lagged income is "preposterously low." We believe that this value appears low only in contrast to lag sums derived from "customary" specifications which employ PDL/levels instead of differences and the Standard rather than the Consolidated Approach. Finally, results similar to those in Table 3 and 4 obtain when the dependent variable is consumption expenditures, both with and without consumer durables.

[16] Here we distinguish between expectations of the future tax consequences of current deficits and the existence of a bequest motive, a distinction first made in Robert Barro's seminal paper (1974).

[17] Actual transfer payments are also paid to agents with low permanent income (from "rich" to "poor"), and to agents who are temporarily liquidity constrained. All three categories of transfer payments, which are not mutually exclusive, imply a positive coefficient for transfer payments. See Kormendi (1983).

[18] For example, Social Security *redistributions* would reduce capital formation, even though the taxes needed to finance the future unfunded Social Security obligations (Social Security "wealth" in the sense of Martin Feldstein, 1974) are virtually fully discounted.

REFERENCES

Aschauer, David A., "Fiscal Policy and Aggregate Demand," *American Economic Review*, March 1985, *75*, 117–27.

Barro, Robert J., "Are Government Bonds Net Wealth?," *Journal of Political Economy*, November/December 1974, *82*, 1094–117.

Barth, James R., Iden, George and Russek, Frank S., "Government Debt, Government Spending and Private Sector Behavior: Comment," *American Economic Review*, December 1986, *76*, 1158–67.

Feldstein, Martin S., "Social Security, Induced Retirement and Capital Accumulation," *Journal of Political Economy*, September/October 1974, *82*, 905–26.

Granger, C. W. J. and Newbold, Paul, "Spurious Regressions in Econometrics," *Journal of Econometrics*, July 1974, *2*, 111–20.

Hall, Robert, "Stochastic Implications of the

Life-Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, December 1978, *86*, 971–87.

Hausman, Jerry A., "Specification Tests in Econometrics," *Econometrica*, November 1978, *46*, 1251–72.

Kormendi, Roger C., "Government Debt, Government Spending, and Private Sector Behavior," *American Economic Review*, December 1983, *73*, 994–1010.

______, "Does Deficit Financing Affect Economic Growth? Cross-Country Evidence," *Studies in Banking and Finance*, forthcoming 1986.

______ **and LaHaye, Laura**, "Cross-Regime Tests of the Permanent Income Hypothesis," working paper, University of Michigan, May 1986.

Mankiw, N. Gregory and Shapiro, Matthew, "Trends, Random Walks and Tests of the Permanent Income Hypothesis," *Journal of Monetary Economics*, September 1985, *16*, 163–74.

Modigliani, Franco and Sterling, Arlie, "Government Debt, Government Spending and Private Sector Behavior: Comment," *American Economic Review*, December 1986, *76*, 1168–79.

Plosser, Charles and Schwert, G. William, "Estimation of a Non-Invertible Moving Average Process: The Case of Overdifferencing," *Journal of Econometrics*, September 1977, *6*, 199–224.

Sargent, Thomas, "Rational Expectations, Econometric Exogeneity, and Consumption," *Journal of Political Economy*, August 1978, *86*, 673–700.

______, *Macroeconomic Theory*, New York: Academic Press, 1979.

The Economics of Price Scissors: Comment

By AKE G. BLOMQVIST *

In a recent paper (1984a), Raaj Kumar Sah and Joseph Stiglitz construct a model which they use for analyzing the relationship between the rate of industrial sector capital accumulation and the terms of trade between industry and agriculture in a closed two-sector economy. They relate some of their analysis to the industrialization debate in the Soviet Union in the 1920's, and particularly to the proposals by Evgeny Preobrazhensky for financing industrial capital accumulation by putting a "squeeze on the farmer through the price scissors" (i.e., by deliberately altering the terms of trade against the agricultural sector). According to their results, increasing the squeeze will lead to increased accumulation, as suggested by Preobrazhensky; however, contrary to the suggestions implicit in the work of authors such as Ashok Mitra (1977) and Michael Lipton (1977), they find that the increase in accumulation will be larger the more price sensitive is the marketed agricultural surplus. Furthermore, they interpret their results as contradicting "Preobrazhensky's second proposition," namely that by turning the terms of trade against the peasants, it is possible to increase the rate of accumulation without reducing the welfare of industrial workers.

The purpose of this comment is to suggest that several of the Sah-Stiglitz conclusions are misleading in the sense that they depend on a model specification which is inconsistent with the spirit of Preobrazhensky's discussion of the policy problem faced by Soviet authorities in the 1920's, and also by policymakers in some present-day less developed countries. If the problem is modeled in a way that corresponds more closely with Preobrazhensky's analysis, not only will Preobrazhensky's second proposition be true, but the suggestion by Mitra and Lipton, that the scope for increasing accumulation through manipulating the terms of trade will be *less* the more price sensitive is the marketed agricultural surplus, will also be correct.

The central relations in the Sah and Stiglitz model are the equations describing the rate of capital accumulation (I) and the supply-demand balance for the agricultural good; following their notation, they can be written as

$$I = N^2\big(Y(\bar{K}/N^2) - w\big), \tag{1}$$

$$N^1 \cdot Q(p, N^1) = N^2 x^2(p, w), \tag{2}$$

where N^1 and N^2 are the number of workers in the agricultural and industrial sectors, respectively, $\bar{K}$ is the given capital stock in industry, Y is the per worker output of industrial goods, w is the industrial wage rate, Q and x^2 are the marketed surplus per agricultural worker and food consumption per industrial worker, respectively, and p is the price of the agricultural good relative to that of the industrial good (the terms of trade).

In establishing the results referred to above, Sah and Stiglitz take N^1 and N^2 as given, and then analyze the effects of changing p which is considered a control variable. Inspection of (2) makes it clear that as long as the marketed surplus responds positively to price and food is a normal good (so that $(\partial x^2/\partial p) < 0$, $(\partial x^2/\partial w) > 0$), a reduction in the relative price of the agricultural good will lead to an excess demand for it. Sah and Stiglitz now assume that this excess demand is eliminated through a reduction in the industrial wage rate w. It is immediately obvious that the reduction in w will raise the rate of accumulation I, and that it must be large enough to actually reduce the real income of industrial workers. As Sah and Stiglitz note, the result that the increase in capital accumulation will be larger when the marketed surplus is highly price responsive simply reflects the fact that the necessary reduction in w must be large when this is the case.

---

*Department of Economics, University of Western Ontario, London, Ontario, Canada, N6A 5C2. I thank Joseph Stiglitz and Victoria Zinde-Walsh for references and comments.

But a process in which the state turns the terms of trade against agriculture by means of reducing the real wages of the urban proletariat does not appear to be what Preobrazhensky had in mind. As Sah and Stiglitz note, in one place he tries to construct a numerical example to demonstrate that increasing accumulation by means of "taxation on the basis of a definite price policy" need not reduce the real income of workers. More generally, the fundamental theme of his chapter on "The Law of Primitive Socialist Accumulation" is that policies must be devised to make the peasants carry more of the burden of industrialization, and the industrial proletariat less.[1] The attacks by contemporary critics on Preobrazhensky's anti-peasant views certainly indicate that this was the understanding at the time.

An alternative and more plausible way of interpreting Preobrazhensky's essay is to see it as a set of proposals implying various forms of indirect taxation of the marketed agricultural surplus. Numerous passages can be cited in support of this interpretation.[2] Such indirect taxation would turn the terms of trade against the peasants, but would be compatible with demand-supply balance for the rural good because it would raise the (tax-inclusive) price to urban consumers.

Formally, indirect taxation of the marketed surplus of agricultural goods can be introduced in the Sah and Stiglitz model by letting the price paid by urban workers (denoted here by $\hat{p}$) be a control variable; the tax rate t on the marketed surplus is then given implicitly by

$$t = (\hat{p} - p)/\hat{p}. \tag{3}$$

The question how this type of indirect tax may be used to raise the rate of accumulation and to redistribute income between farmers and urban workers, has been analyzed in considerable detail in other papers by Sah and Stiglitz (1984b, 1983). For present purposes, the important point to establish is that Preobrazhensky's second proposition is valid when the problem is defined in this way. I do this by showing that a policy of turning the terms of trade against agriculture by taxing the marketed surplus may be used to raise the rate of accumulation even if the real income of urban workers is held constant by an increase in the wage rate sufficient to compensate for the impact of the higher (tax-inclusive) price of food. To demonstrate this, one must first take account of the revenue raised by the tax, so that equation (1) must be written as[3]

---

[1] See in particular Preobrazhensky's distinction between "socialist accumulation" and "primitive socialist accumulation" (p. 84), his statement recognizing that a concern for workers' welfare constitutes a "definite barrier to the tempo of industrialization" (p. 122), and the polemic with Nikolai Bukharin in the Appendix (especially pp. 229–230, 266).

[2] He suggests that in the future, freight charges on the state railway may be used as "a lever of primitive accumulation" (p. 95), as well as profits from state purchases of grain or raw materials for resale to urban consumers or use in state industries (pp. 100, 103). Finally, he advocates indirect taxation through "creation of monopoly prices on the home markets for the products of national industry"; that is, fixing the nominal price of manufactured goods (p. 111). It is in the context of discussing the latter policy that Preobrazhensky suggests that urban workers may be fully compensated by nominal wage increases for the impact of higher industrial prices. In the context of the Sah and Stiglitz model, an attempt at altering the terms of trade through price fixing of this kind would necessarily result in a food shortage unless wages were allowed to fall. Sah and Stiglitz suggest that Preobrazhensky and others did not realize this ("the crucial point missed by Preobrazhensky and others is the constraint generated by the demand and supply balance of the rural good," p. 129). However, the complete passage in Preobrazhensky's essay from which Sah and Stiglitz quote, implies the existence of a second consumer group ("bourgeois and petty-bourgeois consumers") who do not get compensated for the increased price of industrial goods. In these circumstances, raising the nominal price of the manufactured good *could* reduce the equilibrium relative price of the rural good without creating a shortage, even if industrial workers were fully compensated. In Lipton's work, the possibility of a food shortage as a result of a policy of altering the terms of trade against agriculture is discussed very explicitly (for example,

$$(1') \quad I = N^2\big(Y - w + (\hat{p} - p)x^2(\hat{p}, w)\big);$$

this expression is then used to evaluate the total derivative $dI/d\hat{p}$ subject to the constraints

$$(2') \quad N^1 \cdot Q(p, N^1) = N^2 x^2(\hat{p}, w)$$

$$(4) \quad V^2(\hat{p}, w) = \text{constant},$$

where V^2 is the indirect utility function of an urban worker, and p and w are endogenously determined. The required result is then established by showing that $(dI/d\hat{p}) > 0$ at $t = 0$, and that $(dp/d\hat{p}) < 0$. Both of these inequalities are proved in the Appendix where the following expressions are derived:

$$(5) \quad \varepsilon_{p\hat{p}} \equiv (\hat{p}/p)(dp/d\hat{p}) = -\varepsilon_{xp}^{2u}/\varepsilon'_{Qp} < 0,$$

$$(6) \quad dI/d\hat{p} = N^2 x^2 \cdot (-\varepsilon_{p\hat{p}}) \cdot \big(1 - t(1 + \varepsilon'_{Qp})\big),$$

where, following Sah and Stiglitz, ε_{xp}^{2u} is the absolute value of the price elasticity of the compensated demand by urban workers for the rural good, and $\varepsilon'_{Qp} > 0$ is the price elasticity of the marketed surplus.[4] From (6), it is also evident that for positive t, the increase in accumulation for a given relative reduction in the farm price of the agricultural good is *smaller* the larger this price elasticity, contrary to the result of Sah and Stiglitz (1984a, p. 128), but consistent with the suggestions of Lipton and Mitra.

APPENDIX

Differentiating the constraints (2′) and (4) in the text produces

$$(\text{A1}) \quad \begin{bmatrix} N^1(\partial Q/\partial p) & -N^2(\partial x^2/\partial w) \\ 0 & (\partial V^2/\partial w) \end{bmatrix} \times \begin{bmatrix} dp/d\hat{p} \\ dw/d\hat{p} \end{bmatrix} = \begin{bmatrix} N^2(\partial x^2/\partial \hat{p}) \\ -(\partial V^2/\partial \hat{p}) \end{bmatrix}.$$

Using the definitions of price and income elasticities in Sah and Stiglitz, and noting that $(\partial V^2/\partial w) = \lambda$, $(\partial V^2/\partial \hat{p}) = -\lambda x^2$, where λ is the marginal utility of income of an urban worker, this can be rewritten, after some manipulation:

$$(\text{A2}) \quad \begin{bmatrix} \frac{\hat{p}}{p}\varepsilon'_{Qp} & -\frac{1}{x^2} \cdot \alpha_x^2 \varepsilon_{xm}^2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} dp/d\hat{p} \\ dw/d\hat{p} \end{bmatrix} = \begin{bmatrix} -\varepsilon_{xp}^2 \\ x^2 \end{bmatrix},$$

where ε_{xm}^2 and ε_{xp}^2 are the income elasticity and price elasticity (defined to be positive) of the uncompensated demand for the rural good, and $\alpha_x^2 \equiv \hat{p}x^2/w$ is the share of food in an urban worker's budget. Using Cramer's rule and the definition of the price elasticity of compensated demand (again defined to be positive), one obtains

$$(\text{A3}) \quad dp/d\hat{p} = (p/\hat{p})\big(-\varepsilon_{xp}^2 + \alpha_x^2 \varepsilon_{xm}^2\big)/\varepsilon'_{Qp}$$
$$= \big(-\varepsilon_{xp}^{2u}/\varepsilon'_{Qp}\big) \cdot p/\hat{p} < 0,$$

$$(\text{A4}) \quad dw/d\hat{p} = x^2.$$

Totally differentiating (1′) in the text produces

$$(\text{A5}) \quad dI/d\hat{p} = N^2\Big[\big(x^2 + (\hat{p} - p)(\partial x^2/\partial \hat{p})\big) - \big(x^2(dp/d\hat{p})\big) + (dw/d\hat{p})\big(-1 + (\hat{p} - p)(\partial x^2/\partial w)\big)\Big].$$

---

[3] Note that expression (1′) may still be considered valid even if the government does not actually collect the tax $(\hat{p} - p)$, but instead makes food available to urban workers at the price p, via a rationing system. In that case, $\hat{p}$ may be interpreted as a shadow price or black market price of food; the "tax" $(\hat{p} - p)$ then constitutes a transfer to workers through the rationing system and this transfer makes it possible for the government to pay correspondingly lower wages without reducing workers' welfare. With lower wages, accumulation can increase.

[4] As Sah and Stiglitz note, this price elasticity may in principle be negative; however, on the basis of existing empirical evidence they assume it is positive.

Substituting from (A3) and (A4), noting that $(\partial x^2/\partial \hat{p})_{\bar{u}} = (\partial x^2/\partial \hat{p}) + x^2(\partial x^2/\partial w)$, this becomes

$$\text{(A5}'\text{)}\quad dI/d\hat{p} = N^2x^2\big[((\hat{p}-p)/x^2) \times (\partial x^2/\partial \hat{p})_{\bar{u}} - (dp/d\hat{p})\big]$$
$$= N^2x^2\big(t\varepsilon^{2u}_{xp} + (p/\hat{p})\big(\varepsilon^{2u}_{xp}/\varepsilon'_{Qp}\big)\big)$$
$$= N^2x^2\big(\varepsilon^{2u}_{xp}/\varepsilon'_{Qp}\big)\big(t\varepsilon'_{Qp} + (1-t)\big).$$

(A3) and (A5′) can now be used to produce (5) and (6) in the text.

REFERENCES

Lipton, Michael, *Why Poor People Stay Poor: Urban Bias in World Development*, Cambridge: Harvard University Press, 1977.

Mitra, Ashok, *Terms of Trade and Class Relations*, London: Frank Cass, 1977.

Preobrazhensky, Evgeny, *The New Economics*, Oxford: Clarendon Press, 1965.

Sah, Raaj Kumar and Stiglitz, Joseph E., (1984a) "The Economics of Price Scissors," *American Economic Review*, March 1984, *74*, 125–38.

______ **and** ______, (1984b) "Taxation and Pricing of Agricultural and Industrial Goods," NBER Working Paper No. 1338, 1984.

______ **and** ______, "The Town, The Country, and the Rest of the World: Economics of Public Marketing Boards in Developing Economies," mimeo., Department of Economics, Princeton University, 1983.

The Economics of Price Scissors: Comment

By MICHAEL R. CARTER*

In a recent issue of this *Review*, (1984) Raaj Kumar Sah and Joseph Stiglitz model the impact of shifts in the agriculture-industry terms of trade on industrial accumulation and social welfare.[1] The intersectoral terms of trade, they note, was a key issue in the Soviet industrialization debate of the 1920's, and continues to be an important issue in contemporary less developed countries. One seemingly surprising implication of the Sah-Stiglitz model is that a shift in the terms of trade against agriculture increases industrial accumulation despite (or rather, *because of*) a normal agricultural supply response. In their model, a price-induced decline in marketed agricultural surplus requires depression of industrial wages in order to reequilibrate urban food demand with supply at the lower relative food price. The income and substitution effects of these wage and price changes lower urban consumption demand for industrial goods. Together with agriculture's reduced purchasing power, this lower urban demand leads to the price scissors-induced increase in investable surplus out of industrial production.

Thus emerges the strong Sah-Stiglitz result, which they label "Preobrazhensky's First Proposition," that the *more* elastic is agricultural supply response, the more wages must be depressed in equilibrium, and the *more* effectively price scissors work to increase accumulation (see their equation (15)). Apparently, the traditional preoccupation with agricultural supply response, including Preobrazhensky's,[2] is wrong-headed and backwards. But it is argued here that this result obtains only under rather special labor supply and investable surplus assumptions which conform to neither the constraints of the Soviet industrialization debate, nor to most contemporary less developed countries.[3]

I. Labor Supply Assumptions

Sah and Stiglitz assume that labor supply by industrial workers, implicitly measured in productivity units, is fixed by the government and is unresponsive to price changes. This assumption permits wages to be costlessly depressed in their model, something which Preobrazhensky (1926, p. 122) consid-

---

*Assistant Professor. Department of Agricultural Economics, University of Wisconsin, Madison WI 53706. I thank an anonymous referee for helpful comments and Stiglitz for providing a copy of the Sah-Stiglitz paper (1985).

[1] In an unpublished working paper called to the author's attention by a referee, Sah and Stiglitz (1985) extend their earlier analysis to consider points raised in this comment. Their results like those here, weaken the strong implication of their earlier work.

[2] As commentators on the industrialization debate have noted (for example, Nove in his introduction to Preobrazhensky, 1926, and Alexander Erlich, 1960), Preobrazhensky was unable to counter the argument that a strong agricultural supply response would cripple exports, urban food supplies, and, consequently, accumulation. His not quite logically consistent response was to downplay agriculture's ability to withdraw from the market for more than a short run (1926, p. 181; 1927, p. 179), and to say that, in any event, increased accumulation would quickly allow an increased flow of products to agriculture (1926, p. 89). In his final scholarly work, Preobrazhensky seemingly despairs of this and other contradictions, and notes the grave difficulty of constructing socialism in one country (1927, p. 230).

[3] Interestingly, the Sah-Stiglitz analysis, with its emphasis on industrial wages, better portrays the economics of the Stalinist strategy which in some sense resolved the industrialization debate. Michael Ellman (1975) calculates that increased self-exploitation by the urban working class funded the entirety of the huge increase in soviet accumulation between 1929 and 1932. However, the economic mechanisms at work then were distinct from those in the Sah-Stiglitz model. Ellman shows that aggregate prices were not turned against agriculture (despite compulsory low price deliveries from collective farms), and that agriculture did not provide a net investible surplus. As later discussion here makes clear, the accumulating state need not be constrained to lower wages only by the amount required by changes in marketed surplus when agricultural goods can be exported.

ered impossible, at least for political reasons.[4] But, political reasons aside, whatever the form of worker supervision, lower wages reduce incentive for labor effort. Compression of industrial wages as in the Sah-Stiglitz model would therefore in general be expected to reduce industrial productivity and perhaps accumulation. A "they pretend to pay us and we pretend to work" effect, to borrow an expression from contemporary Poland, can be incorporated into their model by specifying industrial labor productivity as a function of the real wage, $Y(w/\bar{p})$, where $\bar{p} = [\alpha_x^2 p + (1 - \alpha_x^2)]$ is a Laspeyres price index of urban consumer goods.

With this change, equation (15) in Sah-Stiglitz becomes

$$(1) \quad dI/dp = N^2(w/p) \times \left[Y'\left(\frac{\varepsilon_{wp} - \alpha_x^2 p/\bar{p}}{\bar{p}} \right) - L^2 \varepsilon_{wp} \right],$$

where $Y' > 0$ captures the impact of the real wage on productivity given fixed labor time, and ε_{wp}, the elasticity of the urban wage with respect to agricultural price, is a positive function of the price elasticity of agricultural marketed surplus. If the real wage declines ($\varepsilon_{wp} \leq \alpha_x^2 p/\bar{p}$), and if the labor disincentive effects are, or become, large enough, then squeezing agriculture with price scissors will reduce accumulation, and will increasingly do so as agricultural supply responsiveness and ε_{wp} increase. While this specification is no substitute for thorough analysis of the intrafirm productivity environment, it does, together with Avinash Dixit's model (see fn. 4), illustrate that Sah-Stiglitz's strong unconditional result is a rather special case.[5]

II. Definition of Investable Surplus

A second assumption behind the Sah-Stiglitz strong version of Preobrazhensky's first proposition is a narrow definition of investable surplus as the difference between domestic production and consumption of industrial products. In the Soviet industrialization debate, however, lagging agricultural exports and the inability to import foreign capital goods were a major preoccupation.[6] Once the direct importance of agricultural exports to accumulation is a recognized, concern about a price scissors-induced agricultural supply strike appears a good deal more rational than the Sah-Stiglitz framework admits.

Sah and Stiglitz's agricultural material balance equation (10) can be rewritten to define net agricultural exports, $E = N^1 Q - N^2 x^2(p, w)$. Investable surplus, equation (11) in Sah-Stiglitz, then becomes

$$(2) \quad I = N^2(Y - WL^2) + (p^s - p)E,$$

where the new price term p^s is given international or shadow price of agriculture goods which is assumed to exceed p. Assuming that there are no quantity constraints in the international market, state export policy can now be used to influence urban food supply independently of marketed surplus. The equilibrium industrial wage can be set independently of p in conjunction with export policy.

---

[4] In Avinash Dixit's (1969) model of the price scissors problem, (real) wages cannot be depressed because of a Lewis-style assumption that labor supply is perfectly elastic at a real wage fixed in terms of food. Contrary to Sah-Stiglitz's strong version of Preobrazhensky's first proposition, the larger is agriculture's supply responses, the *less* effective are price scissors in Dixit's model. In fact, with too large a supply response, price scissors would *reduce* accumulation in the Dixit model (see his equation 9, p. 211).

[5] Sah and Stiglitz (1985) consider the effect of real wages on labor productivity *via* a nutritional linkage, and they subsequently weaken their price scissor proposition in the manner the present analysis suggests. Also analogous to this comment's later discussion, they consider the impact of trade on price scissors. Section VI of their 1984 paper considers a Harris-Todaro labor supply specification which could be used to similarly dampen their price scissor proposition.

[6] For the Soviet Union, which lacked access to foreign capital markets, agricultural exports were one of the few sources of foreign investment goods and technology. Even as agricultural production recovered toward its prewar level in the 1920's, agricultural exports continued to lag, apparently because of the revolution's abolition of the taxes and rents which in the past had led to large involuntary marketings out of peasant agriculture (see Preobrazhensky, 1925, p. 34 ff.).

If in deference to Preobrazhensky the nominal wage is held constant, and exports are determined residually, then differentiation of (2) shows that the price scissors in fact reduce exports and accumulation, the more so the greater is agriculture's supply response, confirming Preobrazhensky's worries (1927, p. 230). Alternatively, exports could be held at some target level, and the wage could be adjusted to equilibriate urban demand to residual agricultural surplus $(Q-E)$ following an agricultural price squeeze. In this case, of course, price scissors increase accumulation. But, within the context of the Sah-Stiglitz framework, there is no reason for the accumulating state to reduce wages by only the amount required by the reduction in marketed surplus. Wage policy could be used to ratify a simultaneous increase in exports which would simulate the "beneficial" effects of a large supply response.[7] The economy with a small actual supply response can always depress wages just as much as the economy with a large supply response, and it can export and accumulate more. The Sah-Stiglitz strong result does not hold when the agricultural good can be exported without constraint.

In summary, Sah and Stiglitz do point out that the urban wage is a critical part of the economics of squeezing agriculture with price scissors. The centrality of agricultural supply response to price scissors can be overemphasized, as Stalin demonstrated by "resolving" the industrialization debate which had become stalemated over the supply response dilemma (see Erlich). But without very special labor supply and investable surplus assumptions, it cannot be concluded that agricultural supply response is nonproblematic.

REFERENCES

Dixit, Avinash K., "Marketable Surplus and Dual Development," *Journal of Economic Theory*, August 1969, *1*, 203–19.

Ellman, Michael, "Did the Agricultural Surplus Provide the Resources for the Increase in Investment in the USSR during the First Five Year Plan?," *Economic Journal*, December 1975, *83*, 844–64.

Erlich, Alexander, *The Soviet Industrialization Debate*, 1924–1928, Cambridge: Harvard University Press, 1960.

Gregory, Paul and Stuart, Robert, *Soviet Economic Structure and Performance*, New York: Harper and Row, 1981.

Preobrazhensky, Eugeny G., (1925) "Economic Notes 1. On the Goods Famine," in Donald Fitzer, ed., *The Crisis of Soviet Industrialization*, White Plains: M. E. Sharpe, 1979.

______, (1926) *The New Economics*, Oxford: Clarendon Press, 1965.

______, (1927) "Economic Equilibrium in the System of the U.S.S.R.," in Donald Fitzer, ed., *The Crisis of Soviet Industrialization*, White Plains: M. E. Sharpe 1979.

Sah, Raaj Kumar and Stiglitz, Joseph E., "The Economics of Price Scissors," *American Economic Review*, March 1984, *74*, 125–38.

______ **and** ______, "Price Scissors and the Structure of the Economy," unpublished paper, May 1985.

---

[7] This is in part what happened in the Soviet Union under Stalin. Agricultural exports increased some 60 percent between 1929 and 1931 as the proportion of grain exported rose from 1 to 14 percent despite almost constant production. Correspondingly, capital goods imports increased sharply (see Paul Gregory and Robert Stuart, 1981, pp. 245 and 285). As indicated in fn. 3, a key to Soviet accumulation was depression of the real urban wage—depression which went beyond what would have been required had agricultural exports not been increased.

The Economics of Price Scissors: Reply

By RAAJ KUMAR SAH AND JOSEPH E. STIGLITZ*

The question of how the funds required for the capital accumulation associated with industrialization are to be raised has a long history. Since the industrial sector is relatively small in the early stages of industrialization, there has been a presumption that funds must primarily come from the agricultural sector. The issue has been how best to extract these funds, what are the effects of such a resource transfer, what is the nature of the limits on the amounts to be extracted, and whether, in this process, it is desirable or necessary to squeeze simultaneously the urban proletariat.

In our 1984 paper, we presented a simple model of a closed socialist economy in which the instruments at the disposal of the government are the *terms of trade* (the relative price of the agricultural good in terms of the industrial good) and the industrial wage, which we believe sheds some light on these questions, particularly in the context of the Soviet industrialization debate. This model is part of a larger research program, in which we have examined similar questions in the context of economies under a variety of trade regimes, with a variety of institutional structures, with and without heterogeneity among individuals within a sector, with and without the disaggregation of commodities produced in each of the two sectors, with and without migration, where the government has at its disposal other policy instruments, where wages in the urban sector are determined in a variety of ways, and in which wages and prices may affect the productivity of workers.[1]

The reason for examining a variety of economic structures is obvious: as the past experience of economic development has shown, there is an enormous diversity in the salient features of developing countries and in the set of policy instruments that a government can or cannot employ. Among our objectives have been (*i*) to examine a range of models emphasizing different sets of structural features, in part to ascertain which features are central to answering the questions at hand, and (*ii*) to develop general reduced-form relationships to describe specific parts of the economy (for instance, to represent the migration mechanism and the urban wage determination), and to show that the central results depend only on the properties of these reduced-form relationships.

The issues raised by Ake Blomqvist and Michael Carter in their comments (1986) are primarily interpretative. They contend that our simple model of a socialist economy is not an appropriate vehicle for examining certain propositions which the Soviet economist Evgeny Preobrazhensky advanced during the Soviet debate. Instead, Carter argues for an open economy specification, and for a particular type of wage-productivity effect. Blomqvist, on the other hand, favors a specification in which the government can set different prices in two sectors (we have referred to the problem of analyzing differential prices, or taxes, in the two sectors as that of *urban-rural pricing*; in contrast to the *price scissors* problem, where both sectors face the same prices). As the commentators are aware, we have analyzed these (and more general specifications) in our 1985a and 1983 papers, respectively. Since it is not possible to present a full analysis within this short reply, we briefly describe some of the qualitative features of price scissors within a simple framework, so that one can see certain implications of alternative assumptions. We then turn to the more doctrinal and interpretative issues: What did Preobrazhensky *really* mean? Which model provides a better description of the Soviet Union at the time of the debate? We remain convinced that our original interpretation was correct.

---

*Yale University, New Haven, CT 06520 and Princeton University, Princeton, NJ 08544, respectively.

[1] See our studies (1983, 1984, 1985a,b,c, 1986). An empirical framework for analyzing urban-rural pricing issues in a multiperson, multicommodity context was developed in Sah (1982); related models have been employed by Avishay Braverman et al. (1984), and others.

I. Price Scissors[2]

A. *An Economy Closed at the Margin*

In an economy facing binding constraints in external trade (and which cannot borrow against the future), a lowering of the price of the rural good—which reduces the supply of rural surplus available to the urban sector—must be accompanied by a lowering of the urban wage to reduce the demand for the rural good, and hence to balance the supply and demand of the rural good. Thus: *The terms of trade cannot be altered if the government cannot (or does not) alter the urban wage.*

If the urban wage is altered then, though urban workers benefit from the lowered price of the rural good, the required reduction in their wage always turns out to be sufficiently large that *the urban workers end up being hurt from a lowering of the terms of trade.* This result does not depend on the direction (or volume) of external trade, or on the nature of the wage-productivity effects (i.e., on the overall effect of a decrease in the price of the rural good as well as in the urban wage on the *net* output of urban workers). If wage-productivity effects are not significant, then a lower urban wage increases the investable surplus; therefore: *A lowering of the terms of trade increases the investable surplus.* Wage-productivity effects are ambiguous in general.[3] One of the special cases under which the above result continues to hold is where the net output of a worker depends on his level of utility;[4] provided a dollar increase in a worker's income yields less than a dollar's increase in his net output (which is what we would typically expect).

B. *Open Economy*

Since in an open economy the demand and supply of the rural good is balanced by adjusting the externally traded quantities, *the terms of trade and the urban wage can be set independently of one another in an open economy.* A lowering of the terms of trade, keeping the urban wage fixed (in terms of the numeraire, industrial good) increases the investable surplus if currently the rural good is not being taxed too heavily (i.e., the domestic price of the rural good is not too low compared to its international price), or if it is being subsidized.[5] This is in part because a typical developing economy is a net exporter of the rural good, and a lower terms of trade implies a higher tax revenue on exports. On the other hand, if the current tax on the rural good is large, then a further reduction in the terms of trade reduces the investable surplus because the revenue loss from the induced reduction in the rural surplus is sufficiently large to offset other potential gains.

The assumption that the urban wage would remain unchanged (in terms of any one of the two goods) is not always compelling. A useful benchmark case is that where it adjusts to keep the welfare of the urban worker unchanged. Then, there exists a critical level of the terms of trade, p^*, which is below the international price, such that: *The investable surplus increases (decreases) with a lowering of the terms of trade if the current terms of trade is above (below)* p^*. Further, since a lower terms of trade hurts rural workers but, by assumption, leaves the welfare of urban workers unchanged, it follows that: *Any price below p^* is Pareto inefficient.*[6]

---

[2] Proofs of the results described in this section are contained in our 1985a paper.

[3] See our 1985a,c papers for a general (reduced-form) representation of wage productivity effects; special cases of this representation are the hypotheses based on labor efficiency, labor quality, and labor turnover. Carter's interpretation that our reduced-form expression represents a nutrition-based link between wage and productivity is, thus, unnecessarily restrictive.

[4] Carter's specification, in which the net output depends on a particular "real wage" can be seen as an approximation of this case. His treatment of this case (see his equation (1)) is not fully satisfactory, however, because the budget shares (used in defining the real wage) remain unaffected even though wage and prices are changing. Moreover, his assertion that the result $dI/dp < 0$ is a rather special case is ungrounded.

[5] Carter claims (based on his equation (2)), but does not show, that a squeeze with price scissors must lower accumulation. This claim is incorrect.

II. Preobrazhensky's Propositions

Here, we ask, what is the appropriate model to serve as a basis for examining two propositions of Preobraz-hensky: (*i*) the state can increase the accumulation by turning the terms of trade against peasants, and (*ii*) the increased accumulation is possible without hurting the industrial workers.

If indeed there is a single model which captures Preobrazhensky's construct, then it is obvious that both Blomqvist and Carter cannot be right because they argue in favor of two quite different models; at least one of them must be wrong. We argue below that both miss the central issue involved. On the other hand, the divergence of views between Blomqvist and Carter, as well as their disagreement with us, can be viewed as a consequence of the ambiguities in Preobrazhensky's writings; after all, Preobrazhensky was not only writing a verbal economics tract, but was also fighting a life and death battle (as subsequent events were to prove) of ideology and polemics.

There is no doubt, however, that the key policy instrument in Preobrazhensky's scheme of primitive socialist accumulation was the terms of trade between the agricultural and the industrial goods; although he did mention a multitude of other instruments such as direct taxation, railway freights, credit policy, printing money, etc. The reason (which holds for many of today's less developed countries as well) is simply that an attempt to alter the terms of trade merely requires the government to change the nominal price of the industrial good; this attempt is more feasible than implementing virtually any other policy instrument. Basing his conclusions on quotations from Preobrazhensky's writings, Alexander Erlich states this point succinctly:

> "The concentration of the whole of the big industries...in the hands of the workers' state increases to an extraordinary extent...the possibility of carrying out...a price policy on the basis of monopoly"...Preobrazhensky did not, to be sure, renounce direct taxation as an instrument of the redistribution of income in favor of socialist industry...Taxation through price, however, was in his view the most effective single device—both because of the "extreme convenience of collection which did not require a penny for a special fiscal apparatus" and for reasons of political expediency. "The way of direct taxation is the most dangerous way, leading to a break with the peasants." [1960, pp. 49–50]

In contrast, a model with urban-rural pricing, which is what Blomqvist argues for, entails not only large administrative costs but it also requires the administrative ability (in which the USSR was clearly lacking during 1917–24) to monitor the "tax border" between the two sectors. Therefore, it should not be surprising that the participants in the Soviet debate as well as those who have subsequently analyzed this debate have viewed the terms of trade (and not urban-rural pricing) as the central instrument in Preobrazhensky's scheme. We also note that Blomqvist's use of an urban-rural pricing model to view some of Michael Lipton's analysis of the Soviet debate might be misleading because a central instrument in Lipton's analysis (1976, pp. 128–29) is, once again, the terms of trade. In another context, Lipton states "The 'scissors' discussion always was, and is, mostly about changing the price of the rural good *relative to the urban good*—not about changing the relative price paid by the two sectors for the same good" (personal communication, 1982, emphasis in the original).

Next consider Carter's comment concerning external trade. It is obvious that what is important is not whether the country trades or not, but whether the economy can or cannot freely increase its trade *at the margin*. A look at the trade figures of the USSR during the relevant period (i.e., between the October Revolution and the beginning of the industrialization debate in 1924) exhibits an extreme decline compared to pre-World War I period. The trade was negligible until 1920; even in 1923, the import volume was less than 11 percent, and the export volume was less than 15 percent, of the respective 1913 levels (see Paul Gregory and Robert Stuart, 1981, p. 267). Though these (or any other trade figures) could, in principle, be consistent with a model with *or* without constraints on external trade, historical facts suggest the former model. Not only was the USSR facing blockades by the allies during much of this period (augmenting the Soviets' fear of "capitalist encirclement") but, even after the Treaty of Rapallo (April 1922), it was facing significant isolation in trade and current business credit (see Maurice Dobb, 1966, ch. 7).

---

[6] Analogous conclusions hold in models with different urban-rural prices. For instance, in the model on which Blomqvist's comments are based, one obtains $dI/dp \gtrless 0$ if $(\hat{p} - p)/p \gtrless 1/\varepsilon^1_{Qp}$. This, in combination with $d\hat{p}/dp < 0$, implies that: *Lowering the rural price below some level is Pareto inefficient.* See our paper (1983) for a discussion of this and other related results.

The importance of these constraints is explicit in Preobrazhensky's views; though he emphasized the *potential* usefulness of external trade to the Soviet state, he clearly noted that "...all kinds of external complications that might not only sever our economic ties with the capitalist countries but will also most effectively retard even that part of socialist construction that is based on the domestic resources of the Republic" (1921, p. 14). Obviously, a fuller understanding of the implications of trade embargoes and interruptions (and of credit constraints) requires an explicit modeling of these phenomena. In stylized aggregate models such as those being discussed here, however, the facts concerning the USSR during 1917–24 (and the understanding that the participants in Soviet debate had of these facts) suggest that a model with constraints on external trade is more plausible than the one without *any* constraints.

III. Concluding Remarks

A virtue of developing a general theoretical framework is that it enables one to isolate the features of the economy which are critical for the issues at hand. Can a manipulation of the terms of trade squeeze the agricultural sector to provide funds for industrial development? If the economy faces constraints on external trade and urban wages cannot adjust, the answer is no. If urban wages do adjust, the answer is yes, but only if the urban workers are made worse off. Further, since wages must adjust more the greater is the sensitivity of agricultural surplus to the terms of trade; the greater this sensitivity, the greater is the increase in the investible surplus from turning the terms of trade against peasants. If the economy does not face trade constraints, or if prices in the two sectors can be set independently, then there is no necessary link between urban wages and the rural prices. Yet, in these cases, we show that it is Pareto inefficient to set the relative price of the agricultural good below some level. That is, a price squeeze of peasants beyond some point is counterproductive, regardless of whether the state cares about the welfare of peasants or not.

REFERENCES

Blomqvist, Ake, "The Economics of Price Scissors: Comment," *American Economic Review*, December 1986, *76*, 1188–91.

Braverman, Avishay, Ahn, Choong Yong and Hammer, Jeffrey S., "Government Deficit Reduction and Alternative Agricultural Pricing Policies in Korea," mimeo., The World Bank, Washington, 1984.

Carter, Michael, "The Economics of Price Scissors: Comment," *American Economic Review*, December 1986, *76*, 1192–94.

Dobb, Maurice, *Soviet Economic Development since 1917*, London: Routledge and Kegan Paul, 1966.

Erlich, Alexander, *The Soviet Industrialization Debate*, Cambridge: Harvard University Press, 1960.

Gregory, Paul R. and Robert C. Stuart, *Soviet Economic Structure and Performance*, New York: Harper and Row, 1981.

Lipton, Michael, *Why Poor People Stay Poor: Urban Bias in World Development*, Cambridge: Harvard University Press, 1976.

Preobrazhensky, Evgeny, (1926) *The New Economics*, Oxford: Clarendon Press, 1965.

______, (1921) "The Outlook for the New Economic Policy," in Donald Fitzer, ed., *The Crisis of Soviet Industrialization*, White Plains: M. E. Sharpe, 1979.

Sah, Raaj Kumar, "A Welfare-based Determination of Urban-Rural Prices," paper presented at the World Bank Conference on Public Economics, Washington, D.C., June 1982.

______ **and Stiglitz, Joseph E.,** "The Town, the Country, and the Rest of World: Economics of Public Marketing Boards in Open Developing Economies," mimeo., Department of Economics, Princeton University, 1983.

______ **and** ______, "The Economics of Price Scissors," *American Economic Review*, March 1984, *74*, 125–38.

______ **and** ______, (1985a) "Price Scissors and the Structure of the Economy," *Quarterly Journal of Economics*, forthcoming.

______ **and** ______, (1985b) "The Taxation and Pricing of Agricultural and Industrial Goods in Developing Economics," in David M. G. Newbery and Nicholas Stern, eds., *The Theory of Taxation for Developing Countries*, Oxford: Oxford University Press, forthcoming 1986.

______ **and** ______, (1985c) "The Social Cost of Labor and Project Evaluation: A General Approach," *Journal of Public Economics*, December 1985, *28*, 135–63.

______ **and** ______, *Peasants versus City-Dwellers: Taxation, Pricing and the Burden of Economic Development*, Oxford: Oxford University Press, forthcoming 1986.

The Production Process in a Competitive Economy: Comment

By ESKANDER ALVI*

In his recent paper, in this *Review*, Samuel Bowles (1985) argues that involuntary unemployment in a competitive capitalist economy is an outcome of the Marxian notion of "class conflict" between the firm (capitalist) and workers. The purpose of this comment is not to question the Marxian framework, but simply to argue that the model and examples in Bowles, though very interesting, do not in any cogent way reflect class conflict.

Let me list the major points made by Bowles in the context of a competitive capitalist economy: 1) involuntary unemployment, leading to "reserve army of the unemployed," is a reflection of class conflict; 2) capitalist technology may be profit maximizing but inefficient; and 3) discrimination among identical workers may be profit maximizing. I will focus primarily on the first observation which is the heart of the paper. The other two can be deduced from the first.

I. Involuntary Unemployment and Class Conflict

Bowles constructs a model of production where total labor input is the product of hours of labor and effort per hour. Since effort is not directly observable, he postulates supervision in the production process. The typical worker (all workers are identical and there is no free-rider problem among workers) is fired if caught shirking. Hence, the effort response of the worker depends on the wage, probability of getting caught when shirking (specified by the employer), and alternative income/utility when unemployed.

The labor market equilibrium that emerges is one where the unemployed worker would rather work than be unemployed at the current wage (i.e., the marginal product exceeds the worker's opportunity cost). Bowles claims that this is the outcome of class conflict between the capitalist and the worker. Such unemployment is indeed painful. However, it reflects the deadweight loss to society from the informational structure rather than class conflict. Unfortunately, Bowles does not say much about this information cost, except

> ...the cost of surveillance is not an information cost at all (or at least a very peculiar one)...if employers know exactly how much work each worker will do once the wage and level of surveillance is selected, would it not be optimal to pay workers according to the amount of work done? It might. [p. 23, fn. 14]

My answer is in the negative. If output is not observed before payment (which must be the implicit assumption), then knowing the worker's *ex ante* optimal effort does not automatically guarantee performance. To illustrate, imagine paying a positive wage to an unknown individual to mow your lawn right before leaving for work (zero supervision). Would you expect to find the lawn mowed after returning from work? In the context of a long-run relationship, it is obvious that supervision requirements can be reduced by tying future payments to present performance. But this is certainly not the point of contention in the Bowles argument.[1] In general, both the firm and workers are worse off, relative to complete information. In the limit, there would be no transaction between the effort-conscious firm and the effort-averse worker in a one-period model without supervision unless one makes the *ad hoc* assumption that some minimum effort input is naturally guaranteed.[2] If we take

---

*Department of Economics, University of Arizona, Tucson, AZ 85721. I thank an anonymous referee for helpful comments.

[1] If anything, such long-term agreements will reduce the loss from incomplete information, and hence reduce the significance of Bowles' argument.

[2] With a guaranteed minimum that is less than potential, it is easy to show that the allocations are a strict

this as a natural starting point in a one-period firm-worker relationship, then supervision merely improves the allocation of resources, albeit inefficiently relative to complete information.

Informational imperfections, particularly asymmetric information, exist on both sides of the market. In such situations, incentive-compatibility may require unemployment, incomplete insurance, and strikes, depending on the type of asymmetry.[3] The relevant question is, why don't firms and workers get together and design a perfect information generating mechanism, since both can be made better off? The answer perhaps lies in the cost of generating such a mechanism. If these costs are increasing and convex in desired mechanisms, such that approximately perfect ones are prohibitively costly, then the society must be content with less than the perfect one, equating marginal benefit (increased welfare to firms and workers via reduced deadweight loss) and marginal cost (the cost of designing the marginally better mechanism).

My next objection is directed at Bowles' attempt to relate this unemployment to a weakening of workers' bargaining power. It is possible to visualize a simple bargaining model where the firm may have an incentive to create excessive unemployment to induce workers to take a lower wage.[4] This is, however, not examined in the paper.[5] In a bargaining model it is only fair to include both the firm's incentive to create excessive unemployment and the workers' ability to collude and strike. However, we must bear in mind that excessive unemployment is also painful to the firm (assumed competitive in the product market) and coupled with the threat of strike, creating excessive unemployment may be only of limited value to the firm. Finally, long-term implicit contracts, enforced via reputation, are more meaningful in situations of bilateral monopoly, and can circumvent such a problem although, in general, imperfectly.

II. Capitalist Technology, Discrimination

I submit that both these claims are an inextricable part of the deadweight loss from incomplete information. Bowles characterizes technology as inefficient because it uses some inputs in the supervision process to induce effort. Clearly it is inefficient relative to complete information; but once we postulate private information, the choice of such a technology cannot be termed inefficient.

Similarly, the claim is made that discrimination among identical workers is also profit maximizing. Correctly, Bowles points out, in the context of a longer time horizon, the existence of promotion ladders, "less now, more later" type of payment institutions to induce effort. This is precisely the point that, in a long-run context, future rewards can be conditioned on present performance, thus reducing the burden of incomplete information. Given the informational structure, however, this can hardly be termed discrimination.

III. Final Remarks

Bowles uses a simple model of production under asymmetric information and interprets the non-market-clearing equilibrium as a reflection of class conflict. I have argued that such unemployment is essentially due to the informational structure.

Before concluding the paper, Bowles makes a distinction between neo-Hobbesian and Marxian models: "The attitude towards work is not, according to this [Marxian] principle, simply a manifestation of human nature [Hobbes], but in part the result of the social institutions in which the production

---

subset of all allocations under complete information. Clearly, supervision is desirable in that it increases the opportunities for trade between the firm and the worker.

[3] See Costas Azariadis (1983), V. V. Chari (1983), Jerry Green and Charles Kahn (1983), Sanford Grossman and Oliver Hart (1981; 1983), and Hart (1983) for a discussion of firms having private information; and myself (1986a,b), Russell Cooper (1983), and Carl Shapiro and Joseph Stiglitz (1984) for discussion of workers having private information.

[4] An example is Martin Osborne (1984).

[5] In all fairness to Bowles, I must admit that this argument is not central to his conclusions. However, he discusses this issue in relation to the Marxian theory of capitalist economy. See p. 25, particularly the second paragraph.

process takes place" (p. 33). I do not argue this point. But one must be able to characterize that social institution which eliminates or reduces, in its own right, the burden of incomplete information. More important, such an institution, to make sense from the point of view of class conflict, must be one that cannot be implemented in a competitive capitalist economy, given its ownership structure.

To my mind, a convincing argument relating unemployment to class conflict should be able to show that involuntary unemployment is a steady-state phenomenon in a competitive capitalist economy (without assumed information problems and externalities) and arises purely due to differences in ownership of capital and labor power.

REFERENCES

Alvi, Eskander, (1986a) "Efficiency-Wage Contracts," Working Paper No. 86-4, University of Arizona, 1986.

______, (1986b) "Information Revelation and Principal-Agent Contracts," mimeo., University of Arizona, 1986.

Azariadis, Costas, "Employment with Asymmetric Information," *Quarterly Journal of Economics*, Suppl. 1983, *98*, 157–72.

Bowles, Samuel, "The Production Process in a Competitive Economy: Walrasian, Neo-Hobbesian, and Marxian Models," *American Economic Review*, March 1985, *75*, 16–36.

Chari, V. V. "Involuntary Unemployment and Implicit Contracts," *Quarterly Journal of Economics*, Suppl. 1983, *98*, 107–22.

Cooper, Russell, "Worker Asymmetric Information and Involuntary Unemployment," Cowles Foundation Discussion Paper No. 671, July 1983.

Green, Jerry and Kahn, Charles M., "Wage-Employment Contracts," *Quarterly Journal of Economics*, Suppl. 1983, *98*, 173–88.

Grossman, Sanford J. and Hart, Oliver D., "Implicit Contracts, Moral Hazard and Unemployment," *American Economic Review Proceedings*, May 1981, *71*, 301–07.

______ **and** ______, "Implicit Contracts under Asymmetric Information," *Quarterly Journal of Economics*, Suppl. 1983, *98*, 123–56.

Hart, Oliver D., "Optimal Labor Contracts under Asymmetric Information: An Introduction," *Review of Economic Studies*, January 1983, *50*, 3–35.

Osborne, Martin, "Capitalist-Worker Conflict and Involuntary Unemployment," *Review of Economic Studies*, January 1984, *51*, 111–127.

Shapiro, Carl and Stiglitz, Joseph, "Equilibrium Unemployment as a Worker Discipline Device," *American Economic Review*, June 1984, *74*, 433–44.

The Production Process in a Competitive Economy: Reply

By SAMUEL BOWLES*

In my 1985 paper, I argue that the difficulty which employers face in controlling the pace and quality of labor input yields a competitive equilibrium pattern of technical choice, involuntary unemployment, and labor market discrimination which is inefficient in a well-defined sense, and that this inefficiency may be attributed at least in part to the capitalist nature of the institutions governing the economy. I further suggest that the extent of work resistance by workers and the difficulty of employers acquiring information about the workers' level of effort are endogenous in the sense that alternative economic institutions would generate different attitudes towards work and different patterns of information cost.

From Eskander Alvi's concluding remarks in his comment (1986), I gather that he agrees that workers' attitudes towards work are endogenous in the above sense. But Alvi also would like to claim that the costs of unemployment (and, by implication, discrimination and nonoptimal technical change) are not attributable to the class structure but rather "reflect the deadweight loss to society from the informational structure" (p. 1200). My point is precisely that the informational structure is not exogenous, but rather reflects in important measure the forms of work resistance, worker solidarity, and commitment to the job fostered by the structure of ownership and control of the labor process.

Quite apart from the endoegeneity of work resistance, however, Alvi's claim that surveillance of the work process (what he calls supervision) "merely improves the allocation of resources" is demonstrably false. Starting from a position defined by the profit-maximizing firm's first-order conditions (my equation (10), p. 24), imagine that the wage is arbitrarily increased by one unit, raising the cost of job loss to the worker and inducing a higher level of work effort. The level of surveillance inputs could then be hypothetically reduced to a level which induces the prior level of work effort. In this new situation the level of output is unchanged and the amount of surveillance input is reduced, all other inputs being unchanged. Thus the profit-maximizing level of surveillance is inefficient; the allocation of resources would be improved by a departure from the firm's profit maximum in favor of a high-wage, low-surveillance labor strategy. The reason is that raising the wage (which is costly to the firm but not resource-using to society) economizes on surveillance inputs which have real opportunity costs.

The cost-minimizing wage is thus suboptimal; or, more pointedly, efficiency wages are inefficient.

Alvi's penultimate paragraph poses the issue of information costs in an especially insightful way—to claim that information costs are endogenous to capitalist institutions, he writes, "one must be able to characterize that social institution which eliminates or reduces... the burden of incomplete information" (p. 1202) and to show that this institution is incompatible with the class structure of capitalism. Exactly. And to claim, as he appears to do, that these information costs are invariant across institutions or that capitalism minimizes these information costs, Alvi would have to give compelling reasons why such information-superior institutions are not feasible. But, he is not able to take up the challenge he has so neatly posed.

Both common sense and some fragmentary empirical evidence (referred to in my earlier paper) suggest that more egalitarian and participatory institutions—whatever their other characteristics—would be likely to reduce the degree of conflict over the pace of work, to reduce work resistance, to enhance the legitimacy and hence effectiveness of systems of control of labor effort and thus reduce information costs associated with the regulation of work.

---

*Department of Economics, University of Massachusetts, Amherst, MA 01004.

But the empirical evidence is hardly conclusive—and perhaps my common sense is the next person's dubious reasoning. The way to find out would be to create an economic and political environment in which experimentation and innovation with new organizational forms flourished and in which we could thus learn more about how egalitarian and democratic economic institutions work. But, as long as credit is rationed according to both the asset position of the debtor and the organizational form of the institution seeking the loan (with a strong—and quite rational—preference of the owners and managers of financial institutions to loan to the wealthy and to capitalist organizational forms), the structure of credit markets is likely to act as a barrier to such innovation and learning.

REFERENCES

Alvi, Eskander, "The Production Process in a Competitive Economy: Comment," *American Economic Review*, December 1986, *76*, 1200–02.

Bowles, Samuel, "The Production Process in a Competitive Economy: Walrasian, Neo-Hobbesian, and Marxian Models," *American Economic Review*, March 1985, *75*, 16–36.

Tobin's *q* and the Structure-Performance Relationship: Comment

By WILLIAM G. SHEPHERD*

The interesting paper by Michael Smirlock, Thomas Gilligan, and William Marshall (this *Review*, December 1984) offers two lessons about market power. The main lesson confirms an "efficient structure" hypothesis: that leading firms' high profits reflect only efficiency (including economies of scale) rather than market power. A secondary lesson favors the use of a q ratio (the firm's market value divided by its asset, or replacement, value) in place of profit rates.

Smirlock et al. reduce the issues to the following regression equation, which I and others have earlier used to analyze firms' profitability (or rents):

$$\begin{aligned}\text{Profitability} = a + b\ &\text{Market Share} \\ &+ c\ \text{Concentration} \\ &+ \text{Other background variables (barriers, growth, etc.)}.\end{aligned}$$

Smirlock et al. theorize that the b coefficient will reflect only superior efficiency, assuming that firms attain high market shares solely by being more efficient. Any monopoly profits will be reflected in (and only in) the c coefficient, which shows the effects of collusion.

Past research has repeatedly shown b to be high compared to c, indicating that market share is the leading element of market power.[1] Using 132 firms drawn from my panel (1972) of large-firm data, Smirlock et al. set out to retest the patterns, using q. They reach flat conclusions: as before, b is high and c is low, but high profitability is now said to reflect only efficiency, not market power. Also, q is concluded to be superior to profit rates for this line of research.

These lessons are emphatic and surprising. But they are not tenable, for the reasons which follow. I restate the logical issues in Section I, focusing on Smirlock et al.'s peculiar assumptions. In Section II, the q ratio's reliability is compared to the rate of profit (called π below). Then in Section III, I present new econometric results and briefly discuss their meaning.

I. Structure and Efficiency

Smirlock et al. assert as "traditional" the eccentric assumption that concentration (as a facilitator of collusion) is the exclusive form of monopoly power.[2] Instead, market dominance is the most direct source of market power. The market power occurs regardless of the origins of the dominance.

Firms may actually obtain market power from several elements of their market positions. One element is the firm's own market share, which embodies direct control over market transactions. A second element is concentration: oligopoly collusion can give a diluted degree of market control. Reflecting this, concentration has consistently shown a weaker association with profitability than has market share. A third element, entry barriers, can also permit a degree of market power and enhanced profitability. Whatever their origins may be, once these elements exist, they can raise prices and profits.

---

*Professor of Economics, University of Massachusetts, Amherst, MA 01002. I am indebted to John Howley for advice and research assistance. I also thank Henry W. de Jong, Ben S. Branch, William J. Adams, and Severin Borenstein, graduate students at the universities of Massachusetts and Michigan, members of the Five-College Seminar in Industrial Organization, participants in a conference at the University of Amsterdam, and a referee for helpful advice.

[1] See my work (1972, 1975, 1979, 1982), Bradley Gale and Ben Branch (1982), and Stephen Martin (1983), among others.

[2] The discussion in this section must be brief, because of limited space. See F. M. Scherer (1980) and my paper (1986) for a fuller coverage.

As the mainstream literature has developed from Alfred Marshall (1920) through Joe Bain (1968), these assumptions have been central, and they have been empirically affirmed.[3] Of course, alternative factors may also play a role, as Harold Demsetz (1973) and others have ingeniously theorized in recent years. Some firms may simply be superior, possibly even for long periods; and/or there may be large technical scale economies.[4] These advantages can raise a firm simultaneously to a high market share and high profit rates. Moreover, in the struggle to gain high market shares, the winner may dissipate some or all of its prospective monopoly profits.

These possibilities, which comprise the "efficient-structure hypothesis," could explain some of the market-share-profitability correlation. But they might instead arise infrequently, relevant only to a small minority of cases. And, as noted, efficient-origin dominance creates market power as surely as does dominance from other sources. The issue is empirical. Persuasive support for Smirlock et al.'s view would require concrete evidence about specific firms, proving that most or all of their high profits reflect superior efficiency.

Smirlock et al. instead simply declare that possibility to be a certainty; they *assume* that efficiency is the sole source of higher market shares, and they also assume that market power is obtained only by collusion. Such assumptions set up an empty test, which does not discriminate between market power and efficiency. The regression only indicates, as before, whether market share is more important than concentration and entry barriers.

II. Profitability and a q Ratio

Profit-rate data have been widely used in the past, though they are known to contain errors, of at least moderate scope.[5] Are q ratios preferable? They may offer certain advantages, but they contain two main categories of defects.[6]

First, q ratios are conceptually debatable. They are a phenomenon of capital markets, not of the firm itself. The ratio's numerator (the firm's current market value) depends on the balance among various investors' expectations about the firm's stream of future profits. This market value is indirect and subjective, based on speculation about future possibilities. It is also unstable, subject to swings fed by changing psychology, rumors, and extraneous factors.

Second, the q ratio is complicated to measure. Smirlock et al. were forced to adopt "intricate" methods, with arbitrary and debatable steps. In the *numerator*, only the common stock price could be used directly. Yet it often shows bizarre movements, as in the 1960's "go-go" stock market era. Smirlock et al. assigned preferred stock an assumed value, other items were assigned book values, and debt was valued by a complex and debatable method. In short, their "market value" numerator is partly book value, and its accuracy is doubtful.

The *denominator* was even more complex and debatable. *Replacement* cost was sought by Smirlock et al., despite decades of literature showing the near-impossibility of defining it clearly or estimating it reliably.[7] They were driven to use artificial constructs to estimate it. The resulting denominator is a hybrid book-and-replacement value figure containing unknown degrees of error and possible bias.

---

[3] They are reinforced by the literature on corporate strategy, which focuses on market share and its market influence. Actual business experience also is permeated with concern about market share, more so than about joint-maximizing or entry barriers.

[4] Yet any important superiority is likely to be eroded quickly by market forces, which hire away superior factors and bid up their prices.

[5] See Bain and George Stigler (1963) for basic reviews. A recent polemic (Franklin Fisher et al., 1983) cited by Smirlock et al. greatly overstates the problems.

[6] Proposals to use q have stressed the advantages and technical features. But no study has directly compared the net advantages of q and π, on conceptual and practical grounds. Note that errors in the accounting measure of profit will also be present in the q ratios.

[7] This is shown by the lengthy debates on public-utility valuation (see for example Alfred Kahn, 1971, and other standard works on utility regulation), as well as the voluminous literature on productivity measurement, using estimates of capital.

Combining these two imperfect parts, the estimated q ratio is more likely to be inferior to π than superior. Indeed, q adds sources of measurement error to the errors already inherent in accounting profits.[8]

III. Econometric Results and Their Meaning

The concepts in Section I can easily be retested, but the interpretation requires care. To the numerous post-1970 econometric studies, Smirlock et al.'s regressions add little but the use of q.[9] Table 1 presents the best models that have been developed, given the available data (both my original data and Smirlock et al.'s).[10] Table 2 focuses on the problems of the Smirlock et al. model.

I note first that π and q are highly correlated with each other, with a simple r^2 of .65.[11] The variables π and q also show similar patterns in the regression results of Table 1. In large part, π and q are merely alternative measures of the same phenomenon.

The regression results in Table 1 present a logical set of easily understood patterns, which closely fit mainstream research.[12] Line 1 is the most logical and complete model for explaining profitability by various measurable elements of market structure. Both market share and (a modification of) concentration are included, in a way which avoids any multicollinearity between the two.[13] Two variables (the logarithm of asset size, and advertising intensity) reflect possible entry barriers, while a sales growth variable filters out the effects of exogenous demand shifts.

Market share emerges as the leading structural variable, with a .24 slope closely in accord with earlier tests. At a market share of 70 percent (the maximum included in these data), the implied rise in profit rate is by 17 points, from 7 to 24 points. In contrast, the "Group" oligopoly variable has little significance. Absolute size has a *negative* coefficient (marginally significant), probably reflecting the problems of X-inefficiency that are endemic in large organizations. Advertising intensity is highly significant, but its maximum increment to profits is only 5 points. The constant term is significant at a reasonable level, suggesting a normal rate of return (in the absence of any market power) at about 7 percent.

A direct overlap of market share and concentration ratios is presented in line 2, for comparison. Only market share's role is changed from line 1, toward a lower value and significance. Since collinearity is present here but not in line 1, the results of line 1 are preferable.

Line 3 adopts general estimates of entry barriers (with dummy variables for "medium" and "high"), in places of size and advertising intensity. These barrier dummies are rough estimates, and so line 3 is less reliable than lines 1 and 2 (as the lower R^2 attests). Line 3 suggests that high barriers can raise profit rates by up to 3 percentage points, but market share's possible 17-point effect is over five times larger than that.

---

[8] To assess the importance of the problems, I requested information from Smirlock et al. on the components of their q data. This information was not supplied.

[9] Moreover, they do not acknowledge that the earlier studies anticipate virtually all of the regression analysis that they offer (apart from the use of q). In particular, market shares and concentration are jointly included in most of the studies noted in fn. 1.

[10] Smirlock et al.'s 132 firms are drawn from my original (1972) 231-firm panel, strictly according to the availability of Compustat data. Meanwhile I have trimmed and improved the data, down to 182 firms. Of those, 117 are on Smirlock et al.'s set of 132. The specific coverage (231 firms, 182, 132, or 117) probably does not introduce bias (except in Smirlock et al.'s results as presented in Table 2 and discussed below).

[11] When a constant term is included, the corrected R^2 is .72. That regression is $q = -1.18(5.29) + .27\pi(17.09)$, with t-values in parentheses. The π coefficient implies that each added 4 points of profit rate will increase market value by the amount of the firm's assets.

[12] See the sources in fn. 1. My 1972, 1975, and 1979 discussions present the logic of the models presented here.

[13] Market share and concentration are highly correlated, because a firm's high market share will also raise the degree of concentration in the industry. To avoid this overlap, one can use concentration minus market share to represent the group of rivals. I first used this group variable in my 1972 paper; see also my later 1975, 1979, and 1982 results.

TABLE 1—REGRESSION ANALYSIS OF PROFITABILITY AND q, 117 LARGE U.S. INDUSTRIAL CORPORATIONS, 1960–69[a]

		Independent Variables									
Regression Number	Dependent Variable	Constant Term	Market Share M	Concentration C	Group $(C-M)$	Log of Asset Size	Advertising-Sales Ratio	Medium Barriers	High Barriers	Growth	Corrected R^2 (F value)
1	Profit Rate (π)	+7.69	+0.24		+0.04	−0.56	+0.26			+0.51	0.573
		(3.21)	(8.42)		(1.55)	(1.79)	(3.98)			(2.09)	(32.18)
1q	q Ratio	+1.09	+0.06		+0.01	−0.19	+0.09			+0.27	0.558
		(1.42)	(7.05)		(0.79)	(1.94)	(4.09)			(3.48)	(30.28)
2	π	+7.69	+0.20	+0.04		−0.56	+0.26			+0.51	0.573
		(3.21)	(6.27)	(1.55)		(1.79)	(3.98)			(2.09)	(32.18)
2q	q	+1.09	+0.06	+0.01		−0.19	+0.09			+0.27	0.558
		(1.42)	(5.64)	(0.79)		(1.94)	(4.09)			(3.48)	(30.28)
3	π	+3.77	+0.23	+0.02				+1.76	+2.99	+0.51	0.516
		(2.38)	(6.71)	(0.74)				(1.88)	(2.32)	(1.95)	(25.70)
3q	q	−0.21	+0.07	+0.004				+0.26	+0.59	+0.27	0.475
		(0.41)	(5.96)	(0.41)				(0.86)	(1.40)	(3.21)	(21.96)
4	π	+9.28	+0.23			−0.52	+0.28			+0.53	0.568
		(4.26)	(8.26)			(1.66)	(4.20)			(2.16)	(39.14)
4q	q	+1.34	+0.06			−0.19	+0.09			+0.28	0.559
		(1.95)	(7.13)			(1.89)	(4.25)			(3.53)	(37.82)
5	π	+7.49		+0.12		−0.67	+0.31			+0.96	0.428
		(2.70)		(4.91)		(1.85)	(4.16)			(3.52)	(22.67)
5q	q	+1.03		+0.03		−0.22	+0.10			+0.40	0.436
		(1.19)		(3.90)		(2.01)	(4.30)			(4.73)	(23.44)

[a] The t-ratios are shown in parentheses. Profitability is the rate of return on book equity.

TABLE 2—REGRESSION ANALYSIS USING CHANGE IN MARKET SHARE INSTEAD OF GROWTH, LARGE U.S. CORPORATIONS, 1960–69[a]

		Independent Variables						
Regression Number	Dependent Variable	Constant Term	Market Share	Concentration	Medium Barrier	High Barrier	Change in Market Share	Corrected R^2 (F value)
117-Firm Panel (as in Table 1)								
1	π	+3.95	+0.24	+0.03	+1.68	+2.92	+0.11	0.513
		(2.48)	(7.25)	(0.86)	(1.78)	(2.35)	(1.78)	(25.44)
1q	q	−0.10	+0.07	+0.01	+0.21	+0.54	+0.07	0.477
		(0.18)	(6.67)	(0.64)	(0.69)	(1.29)	(3.28)	(22.15)
Smirlock et al. Panel of 132 Firms								
2	π	−1.14	+0.18	+0.05	+0.70	−0.14	+5.38	–
		(0.36)	(5.27)	(1.48)	(0.69)	(0.10)	(2.07)	(11.72)
2q	q	−2.60	+0.06	+0.01	+0.26	+0.38	+2.67	–
		(3.51)	(3.51)	(1.05)	(0.90)	(0.95)	(4.50)	(17.06)

[a] The t-ratios are shown in parentheses. A dash denotes not reported.

Lines 4 and 5 fill out the coverage by showing the separate partial roles of market share and concentration. Concentration is highly significant (in line 5), but less so than is market share. Evidently, it is the firm's own market position that matters the most, rather than the firm's being in a concentrated market.

Now I can turn to Smirlock et al.'s q ratio. To assess its contribution, consider alternate lines 1q through 5q. The results are similar; market share's primacy is affirmed. But high barriers are no longer significant (in line $3q$). Moreover the q intercept in line $3q$ is anomalous, at a *negative* .21. It should be near a *positive* value of 1.0; a zero market position should hold a firm's market value close to its asset value, but never drive it down to negative values. The intercept is close to 1.0 in lines $1q$, $2q$, $4q$, and $5q$ (though in line $4q$ it is rather high at 1.34).

Therefore the Smirlock et al. version of a q ratio gives similar patterns to π, though with slightly looser fits. Apparently, investors' actions are correlated with accounting profit rates, and both are associated with market shares and other structural elements. There is no apparent gain from using a q ratio.

Even so, the q ratio tests do fit other evidence about market share's role. The .06 coefficient in line $1q$ suggests that a 17-point rise in market share will double the q ratio. A dominant firm (at a 70 percent market share) has a predicted market value five times as large as its asset value. Such steep capital gains are a powerful reward (and incentive) in the endless struggles for market share. They help explain the intensity of the efforts and the primacy of market share as the operative variable.

All this is in line with mainstream research. Without q, Smirlock et al.'s tests add nothing and, in fact, tend to create confusion. The trouble occurs because they 1) used both market share and bare concentration together, and 2) replaced the growth variable with the estimated change in market share (ΔM).

Table 2 reproduces their coefficients (mainly in line $2q$), along with parallel results from my 117 firms drawn from their 132. Perhaps because it is a less reliable variable, ΔM causes bizarre results in line $2q$. Smirlock et al.'s q intercept is a *negative* 2.60, and market share's coefficient drops sharply below its well-established range. Barriers are not significant. The whole equation's significance is much reduced. The retesting using my 117 firms (lines 1 and $1q$) shows much less distortion, though the q intercept is aberrant at a negative .10. Barriers are significant in the profitability regression, but not with q. Apparently some technical feature of ΔM or Smirlock et al.'s procedures caused peculiar results. Line 1 of Table 1 is the more appropriate basis for drawing any inferences about market power vs. efficiency.

But it does *not* distinguish between the two, despite Smirlock et al.'s claim. Other research indicates that technical scale economies are only a minor cause of the market-share-profitability association.[14] Superior management or innovation may be at work, but their importance is an open question.[15] Smirlock et al.'s test simply does not provide evidence about that. In any event, dominance can provide market power greater than occurs from collusion.

Therefore, the substantive issue of market power and efficiency remains open. As for the q ratio, it might supplement the careful use of profit rates, but it is far from a miracle measure.

---

[14] Scale economies in these firms rarely extend beyond 10 percent market shares, and in many cases they are below 5 percent (Scherer). In a direct test of 85 firms drawn from my large-firm panel (including all those with market shares over 20 percent), scale economies accounted for only 5 to 15 percent of supranormal profits (see my 1982 paper).

[15] Scherer's thorough review (pp. 280–95) suggests a minor role.

REFERENCES

Bain, Joe S., *Industrial Organization*, rev. ed., New York: Wiley & Sons, 1968.

Demsetz, Harold, "Industry Structure, Market Rivalry, and Public Policy," *Journal of Law and Economics*, April 1973, *16*, 1–9.

Fisher, Franklin M., McGowan, John J. and Greenwood, Joen E., *Folded, Spindled and Mutilated: Economic Analysis and U.S. vs. IBM*, Cambridge: MIT Press, 1983.

Gale, Bradley T. and Branch, Ben S., "Concentration versus Market Share: Which Determines Performance and Why Does it Matter?," *Antitrust Bulletin*, Spring 1982, *27*, 83–105.

Kahn, Alfred E., *The Economics of Regulation*, 2 vols., New York: Wiley & Sons, 1971.

Marshall, Alfred, *Principles of Economics*, 8th ed., London: Macmillan, 1920.

Martin, Stephen, *Market, Firm, and Economic Performance*, Monograph Series in Economics and Finance, New York University, 1983.

Scherer, F. M., *Industrial Market Structure and Economic Performance*, 2d. ed., Boston: Houghton Mifflin, 1980.

Shepherd, William G., "The Elements of Market Structure," *Review of Economics and Statistics*, February 1972, *54*, 25–37.

______, *The Treatment of Market Power*, New York: Columbia University Press, 1975.

______, *The Economics of Industrial Organization*, Englewood Cliffs: Prentice-Hall, 1979.

______, "Monopoly Profits and Economies of Scale," in John Craven, ed., *Industrial Organization, Antitrust, and Public Policy*, Boston: Kluwer Nijhof, 1982.

______, "On the Core Concepts of Industrial Economics," in Henry W. de Jong and W. G. Shepherd, eds., *Mainstreams in Industrial Organization*, Amsterdam: Martinus Nijhof, 1986.

Smirlock, Michael, Gilligan, Thomas and Marshall, William, "Tobin's q and the Structure-Performance Relationship," *American Economic Review*, December 1984, *74*, 1051–60.

Stigler, George, *Capital and Rates of Return in Manufacturing*, Princeton: Princeton University Press, 1963.

Tobin's *q* and the Structure-Performance Relationship: Reply

By MICHAEL SMIRLOCK, THOMAS GILLIGAN, AND WILLIAM MARSHALL*

In our earlier paper (1984), we attempted to test two alternative models of the sources of economic profits in major industrial corporations. The first model, which we referred to as the traditional hypothesis, is the classical structure-conduct-performance theory pioneered by Joe Bain (1951) and others. The traditional hypothesis states that the persistence of economic profits is an indication of allocative distortions and the result of elements of market structure that facilitate collusion or otherwise retard the competitive process. The second theory, which we referred to as the efficient-structure hypothesis, is the differential efficiency theory posited by Harold Demsetz (1973), Sam Peltzman (1977) and others. The efficient-structure hypothesis argues that the persistence of economic profits is an indication of the relative efficiency of some firms and results from the possession of scarce skills or factors of production.

The vehicle for the test of these two theories was an ordinary least squares regression of a firm's profitability (as measured by Tobin's q) against the elements of market structure associated with the traditional hypothesis (concentration, entry barriers, and growth rates) and a variable that proxies a firm's relative efficiency (the firm's market share). The results of the estimation provided strong support for the efficient-structure hypothesis but only weak support for the traditional hypothesis.

In his comment on our paper, William Shepherd has two criticisms of our analysis. He also conducts an econometric exercise. As discussed below, we find very little in his comment which either adds to or modifies the findings of our earlier paper.

I. Interpretation of *q* and Market Share

Shepherd's first criticism involves the use of market share as *solely* a proxy for a firm's relative productive efficiency. The differential efficiency of firms can result in distributions of market shares that deviate from those characterizing industries where all firms enjoy equal access to a homogeneous technology. Firms with superior efficiency through access to some Ricardian factor can experience higher market shares. Indeed, it is easy to show that under quite plausible conditions (that increases in the Ricardian factor reduce marginal costs), the market share of the relatively efficient firm will be higher than its less fortunate competitor. Shepherd appears to agree with this point. He claims, however, that market share, like concentration or entry barriers, is an element of market structure which "embodies [the firm's] direct control over market transactions" (p. 1205). Market share, therefore, confers market power and motivates the observed correlation between it and firm profitability. Using market share as a proxy for a firm's relative efficiency cannot, therefore, help differentiate the competing hypotheses.

Shepherd argues that firms may obtain market power (and the opportunity for extra profitability) from their market positions and that the firm's market share is one element of its position. In other words, market share and market power are directly correlated. However, even within the structure-conduct-performance paradigm, the relationship between market share and market power is ambiguous. Shepherd (1972, p. 25) acknowledges the predictive impotence of the structure-conduct-performance paradigm regarding the relationship between market share and profitability. Outside the structure-conduct-performance paradigm, the argument that there is a positive relationship between market share and market power becomes even more tenuous. The extreme case, of course, is a contestable market model where even a firm with the maximum market share has no power to set prices above competitive levels. Many more intermediate models can be constructed to demonstrate the fragile relationship between market power and market share. A clearer specification of the sources of market power conferred by market share is required to justify Shepherd's assertion. To assume that such a relationship exists is simply not tenable.

---

*Assistant Professor of Finance, Wharton School, University of Pennsylvania, Philadelphia PA 19104; Assistant Professor of Economics, California Institute of Technology, Pasadena CA 91125; and Associate Director, Financial Strategies Group, Goldman, Sachs and Co., New York, NY 10004, respectively.

The second criticism raised by Shepherd is concerned with the calculation of q and its validity as a measure of firm rents. Shepherd describes the market value of the firm used in calculating q as "fundamentally indirect and subjective" and based on an implied capricious nature of market participants. Indeed he describes the stock market as often showing "bizarre movements" in prices. Despite Shepherd's assertions to the contrary, the vast majority of empirical evidence is consistent with the efficient market doctrine and supports the view that the market is the best method for assessing a firm's value. Indeed, recent research has shown that stock market prices *lead* changes in real activity (and, by inference, future firm profitability).

Any measure of firm rents must consider both the future stream of firm profits and the risk involved in that profit stream. q does this, while simple accounting profits do not. The benefits from using a measure of future profits and firm risk, such as q, justifies (at least to some extent) the assumptions used in calculating such measures. It is difficult to see why current profits would be preferable, save for the fact that they are easily computed.

II. Analysis of the Shepherd Empirical Findings

Shepherd also presents his own empirical results, which look very similar to our own. He uses market share (MS) and concentration (C) in the same regression (with q or π as the dependent variable) and finds that market share is positive and significant while in no case is the concentration variable significant. We find Shepherd's empirical analysis, however, both confusing and flawed.

First, Shepherd argues that any regression estimation of firm profitability that includes both MS and C suffers from collinearity. Footnote 11 in our original paper deals extensively with these issues, stating that various specifications yield similar results and that degrading collinearity is not a problem. Despite this, Shepherd is unconvinced. He claims to avoid collinearity between the MS and C by using the variables MS as G, where G is defined as $(C - MS)$ in his equations (1). Presumably, Shepherd believes G provides a correction for collinearity because it is uncorrelated with MS.

To believe this simple transformation corrects for collinearity is incorrect, as any elementary econometrics text will show. This should have been obvious from comparing his equations 1 to equations 2, the latter of which include C and MS, but not G. All the summary statistics for 1 and 2 are identical and all the coefficients and t-statistics on the independent variables (except MS) are the same. The coefficient and t-statistics on C in equation 2 is, of course, the same as on G in equation 1. The coefficient on MS in equation 2 is simply the coefficient of MS less the coefficient on G in equation 1. Shepherd's equations 1 and 2 are completely redundant and provide no correction for collinearity nor any additional insight. To argue that "since collinearity is present (in equation 2) but not in line 1, the results of line 1 are preferable" is simply wrong.

Second, Shepherd claims that the higher R^2's associated with π over q suggest "no apparent gain from using a q ratio" (p.1209). To compare R^2s across estimation with different *dependent* variables is not appropriate and quite misleading (even when there is virtually no difference, as is the case here). If one wants to maximize R^2, just choose a dependent variable with very little variation and let the intercept provide your explanatory power (without knowing the variance of q and π in Shepherd's sample, we can only note the significant intercept terms in his π, but not his q, equations).

Third, Shepherd is, in our opinion, overzealous in his emphasis on the sign and magnitude of the intercept term. Unless there are any variable values close to zero, it is difficult to give much weight to the intercept term since the standard error around it as a predictive value will be large.

As a result of the above considerations, we do not believe Shepherd's criticisms of our work are valid, nor do they provide theoretical or empirical evidence that firms with high market share have market power.

REFERENCES

Bain, Joe S., "Relation of Profit-Rate to Industry Concentration: American Manufacturing, 1936–1940," *Quarterly Journal of Economics*, August 1951, *65*, 293–324.

Demsetz, Harold, "Industry Structure, Market Rivalry, and Public Policy," *Journal of Law and Economics*, October 1973, *16*, 1–9.

Peltzman, Sam, "The Gains and Losses from Industrial Concentration," *Journal of Law and Economics*, October 1977, *20*, 229–63.

Shepherd, William G., "The Elements of Market Structure," *Review of Economics and Statistics*, February 1972, *54*, 25–37.

______, "Tobin's *q* and the Structure-Performance Relationship: Comment," *American Economic Review*, December 1986, *76*, 1205–10.

Smirlock, Michael, Gilligan, Thomas and Marshall, William, "Tobin's *q* and the Structure-Performance Relationship," *American Economic Review*, December 1984, *74*, 1050–60.

Pechman's Tax Incidence Study: A Note on the Data

By EDGAR K. BROWNING*

In his recent study (1985), Joseph Pechman presents estimates of tax incidence by income classes for several recent years. This study basically updates the earlier work by Pechman and Benjamin Okner (1974), which developed estimates for 1966, using the same methodology but applying it to more recent data to generate estimates of tax burdens by income class for 1970, 1975, 1980, and 1985. This note points out some peculiarities in the underlying data which suggest there are serious problems with the estimates for the later years.

Pechman's major results are displayed in Table 1. Panel A of the table gives average tax rates for each major tax by household income deciles in 1966, and panel B presents the estimates for 1980, both for incidence variant 1*c*.[1] (Ignore the last column in panel B for the moment.) As can be seen, the combined tax rate for the top decile is estimated to fall between 1966 and 1980 while it rises for the lowest decile. The ratio of the tax rate for the top decile to the tax rate for the bottom decile declines from 1.8 in 1966 to 1.3 in 1980, a significant decrease in progressivity.

Pechman explains the diminished progressivity by noting that the relative importance of various taxes has changed since 1966: "The tax system became less progressive... primarily because the corporation income tax and the property tax declined in importance while more emphasis was placed on the payroll tax" (p. 10). Although it is true that there has been a change in the relative contribution of various taxes since 1966, this change does not account for the decline in progressivity shown by the 1980 estimates. This is easily demonstrated by scaling up or down each decile's 1966 tax rate due to each tax in proportion to the change in the overall tax rate for that tax between 1966 and 1980. (For example, since property taxes had fallen from 3.0 percent of total income in 1966 to 2.0 percent in 1980, we would take two-thirds of the 1966 rates, i.e., 1.4 percent for the lowest decile, as an estimate of the 1980 tax rate due to the diminished importance of the property tax.) The last column of panel B shows the results of these calculations, summed over all taxes. These figures suggest that if the only change since 1966 had been due to the change in the relative importance of various taxes, the tax system in 1980 would have been about as progressive as it was in 1966.

Careful inspection of Table 1 clearly indicates that something else besides a change in the relative importance of taxes was responsible for the difference in results between 1966 and 1980. Note in particular the change for payroll taxes (principally the Social Security tax). In 1966, the payroll tax rate was 2.6 percent for the lowest decile, but in 1980 it is shown as having risen to 8.8 percent, an increase of 223 percent even though payroll taxes as a percent of total income rose by a scant 32 percent over this period. Even more puzzling is how the rate structure of the payroll taxes changed. In 1966, payroll taxes are shown as progressive up through the fifth decile and regressive at higher deciles, but in 1980 they are shown as regressive throughout the entire income distribution.

At this point it will be helpful to explain why other studies (as well as Pechman's 1966 estimates) have found payroll taxes to be progressive at the bottom of the income distribution and regressive at the top.[2] The

---

*Department of Economics, Texas A&M University, College Station, TX 77843.

[1] Eight different sets of incidence assumptions were used in the Pechman study. Variant 1*c* tends to embody competitive assumptions regarding the response of the economy to taxes.

[2] See, for example, my study with William Johnson (1979) or Richard Musgrave, Karl Case, and Herman Leonard (1974).

TABLE 1—AVERAGE TAX RATES BY DECILES OF ADJUSTED FAMILY INCOME

Population Decile	Individual Income Taxes	Corporation Income Tax	Property Tax	Sales and Excise Taxes	Payroll Taxes	Personal Property and Motor Vehicle Taxes	Total Taxes	Total Taxes (Adjusted)
A. 1966								
1	1.1	1.7	2.1	8.9	2.6	0.4	16.8	
2	2.3	2.1	2.6	7.8	3.8	0.4	18.9	
3	4.0	2.2	2.6	7.1	5.4	0.4	21.7	
4	5.4	1.9	2.1	6.7	6.1	0.4	22.6	
5	6.3	1.7	1.8	6.4	6.3	0.3	22.8	
6	7.0	1.5	1.6	6.1	6.2	0.3	22.7	
7	7.5	1.6	1.7	5.7	5.8	0.3	22.7	
8	8.3	1.8	1.8	5.5	5.4	0.3	23.1	
9	8.8	2.2	2.2	5.0	4.8	0.3	23.3	
10	11.4	8.1	5.1	3.2	2.2	0.2	30.1	
All Deciles	8.5	3.9	3.0	5.1	4.4	0.3	25.2	
B. 1980								
1	2.1	0.6	0.7	8.4	8.8	0.1	20.6	15.8
2	3.8	0.5	0.7	7.0	8.3	0.1	20.4	17.3
3	5.3	0.7	0.9	5.9	7.6	0.1	20.6	21.2
4	6.8	0.8	0.9	5.5	7.5	0.2	21.9	23.2
5	8.0	0.9	1.0	5.1	7.5	0.2	22.8	23.8
6	8.8	0.9	1.2	4.9	7.2	0.3	23.3	24.2
7	9.5	1.1	1.2	4.6	6.8	0.3	23.6	24.0
8	11.0	1.2	1.3	4.5	6.7	0.3	25.0	24.5
9	12.1	1.6	1.7	3.9	6.1	0.2	25.7	24.5
10	13.4	5.0	3.5	2.1	3.2	0.1	27.3	28.7
All Deciles	10.8	2.5	2.0	4.0	5.8	0.2	25.3	

Source: Pechman, Tables 4-10 and A-1, except for the Total Taxes (Adjusted) column which gives calculations by the author.

Social Security payroll tax[3] is levied at a flat rate on labor income up to a ceiling amount. For families with earnings below the ceiling, the tax is proportional to labor income but it need not be proportional to total income. For example, if the tax is 10 percent of earnings up to the ceiling, and a family receives 40 percent of its total income as earnings, the tax will be 4 percent of total income. Low-income families typically receive a small percentage of their total incomes in the form of earnings since government transfers are a large proportion of the total. (The before-tax income concept used in incidence studies generally counts transfers as part of before-tax income to be faithful to the Haig-Simons definition of income.) As we move up the income distribution, however, the proportion of income received as transfers typically falls, so the proportion of income received as earnings rises. This relationship implies the payroll tax rate will be increasing at low-income levels. At high-income levels, families increasingly have earnings above the ceiling and a larger share of income is in the form of capital income, so the payroll tax rate tends to decline.

This discussion suggests why the proportion of income received as labor income, capital income, and transfer income by each income class has important implications for the estimated tax rates relative to total income. In an effort to see whether changes in the composition of income could explain the changes in Pechman's estimated tax burdens between 1966 and 1980, I obtained detailed background tables giving data not included

---

[3] The payroll tax category includes, in addition to the Social Security payroll tax, unemployment insurance taxes. These taxes, which are structured similarly to the Social Security tax but at much lower rates and ceilings, are ignored in the discussion here.

TABLE 2—SOURCE OF INCOME BY DECILES OF ADJUSTED FAMILY INCOME
(Percentages)

	1966			1980		
Decile	Labor	Capital	Transfer	Labor	Capital	Transfer
1	30.7	15.9	53.3	82.7	7.9	9.4
2	42.4	19.7	38.0	74.5	8.2	17.3
3	59.4	20.1	20.6	69.4	11.1	19.5
4	71.0	16.5	12.5	70.1	12.3	17.6
5	75.0	15.0	10.0	71.0	13.5	15.5
6	78.0	13.8	8.2	71.2	15.3	13.5
7	78.2	14.4	7.3	68.9	16.7	14.4
8	78.2	14.9	6.9	70.0	17.5	12.5
9	75.4	18.0	6.6	66.4	22.6	11.0
10	51.5	42.9	5.6	44.2	48.5	7.3

Source: Calculated from Tables E-21, E-32, and E-31, provided by Pechman.

in the final monograph from Pechman. From these tables I calculated the proportions of income from each source for each decile in 1966 and 1980; the results are shown in Table 2. According to these data, there has been a phenomenal change in the composition of income over the period, especially for the lower-income classes. The share of labor income is shown to increase greatly for the lower deciles, the share of capital income has fallen sharply, and the share of transfer income has fallen even more dramatically. In addition, note that the share of income received as earnings was rising up to the seventh decile in 1966, but in 1980 the labor share tends to decline throughout the entire income distribution.

These changes in the composition of income received by deciles make it easy to understand why the estimated tax rates differ as they do between 1966 and 1980. With the labor share nearly tripling for the lowest decile, the increase in its payroll tax rate is easily understandable. Moreover, the regressivity of the payroll tax in 1980 can be seen to be the result of the labor share declining throughout the income distribution, rather than rising at the bottom as it did in 1966. Similarly, the very sharp reductions in corporate and property tax rates for the lower deciles are simply a reflection of the diminished capital shares at these income levels. Thus, changes in the estimated composition of income—and not changes in the relative importance of various taxes—appear to be largely responsible for the different results obtained in 1980.

This brings us to the question of whether the composition of income really did change as drastically as Pechman's data indicate. Let me focus primarily on the transfer share since the smaller transfer share for the lower deciles is largely responsible for their higher labor shares.[4] Note that the transfer share for the lowest decile is shown to fall from 53.3 percent in 1966 to 9.4 percent in 1980 in Table 2. There are several reasons for doubting that this actually occurred. First, between 1966 and 1980, transfers as a percentage of net national product more than doubled, rising from 5 to 11 percent. Since the most rapidly increasing transfers were those targeted on the poor, one would expect the transfer share of lower-income classes to rise, not fall, over this period. Second, Pechman's 1966 data do not include in-kind transfers but his 1980 data do.[5] Since in-kind transfers totalled more than $75 billion in

---

[4] Although space limitations preclude a discussion of all the peculiarities in the data, note should be taken of the ratios of capital to labor income in the various deciles. For example, for the lowest decile in 1966, capital income was 52 percent of labor income, but in 1980 it is only 9.5 percent. Even more surprising, the 1970 data (not shown here) have capital income 131 percent of labor income for the lowest decile, but the fraction falls to 10 percent just five years later in 1975.

[5] This is not stated in the monograph, but Pechman pointed this out in private correspondence.

TABLE 3—SOURCE OF INCOME BY DECILES

Decile	Labor	Capital	Transfers
A. 1976			
1	26.0	14.3	59.7
2	34.3	10.9	54.8
3	48.0	11.6	40.5
4	61.0	11.5	27.5
5	70.9	10.7	18.4
6	77.2	10.2	12.6
7	80.7	9.9	9.4
8	81.5	10.6	7.9
9	79.8	13.7	6.5
10	55.7	40.0	4.3
B. 1980		Other[a]	
1	13.1	9.0	77.8
2	34.2	10.8	55.0
3	55.7	12.0	32.3
4	67.7	11.0	21.2
5	75.1	10.0	15.0
6	83.4	7.0	9.5
7	85.7	7.7	6.6
8	87.7	6.3	6.0
9	89.4	6.4	4.2
10	86.4	10.4	3.2

Source: Panel A: myself and Johnson, Table 5; Panel B unpublished data from the *Current Population Survey.*

[a] Includes interest, dividends, rents and royalties, estates and trusts, private pensions, alimony and child support, private financial assistance, and income not included elsewhere.

1980, and are more concentrated on low-income families than cash transfers are, the inclusion of in-kind transfers in the 1980 data should have reinforced the tendency for the transfer share to have risen for the lower-income classes.

Finally, there are also some independent estimates of the importance of transfers by income level that cast doubt on Pechman's data. Panel A, Table 3, gives estimates of decile income shares in 1976 from the Browning-Johnson tax incidence study. (It should be mentioned that our study used a comprehensive measure of income much like that employed in the Pechman study.) Transfers are estimated to be nearly 60 percent of the income of the bottom decile, and the transfer share declines as we move up the income distribution. (Note that in Pechman's 1980 data, the transfer share rises up through the third decile, and is higher than the bottom decile for every income class except the top one.) Since transfers continued to rise rapidly in the late 1970's, one would expect the transfer share to be even higher by 1980.

Panel B, Table 3, gives estimates of income shares by household money income deciles for 1980 from the Census Bureau's *Current Population Survey.*[6] Although the Census Bureau's measure of income is narrower than Pechman's, and includes only cash income (and cash transfers), the results are in general accord with the Browning-Johnson estimates. The transfer share is 77.8 percent for the lowest decile and declines throughout the income distribution. Use of a broader income definition would tend to reduce the transfer share, but inclusion of in-kind transfers would tend to increase it, so the net effect is not clear.[7,8]

These factors lead me to believe that Pechman's data for 1980 greatly underestimate the transfer shares of lower-income classes, and correspondingly overestimate the labor shares. If this is so, it has important implications for the estimated tax rates, both overall and for specific taxes. The exact implications will depend, of course, on the tax incidence assumptions that are used. For the competitive (Pechman's variant 1*c*) assumptions, a higher transfer share for any decile implies lower tax rates under individual income, corporation income, property, and payroll taxes, because these taxes are allocated to labor income, capital income, or both, and a higher transfer share implies lower labor and capital income shares. For example, if the transfer share for the lowest decile in 1980 was 65 percent instead of 9.4 percent, while the tax rates of these taxes applying to labor and capital income remain unchanged, the combined tax rate due to these taxes would fall from 12.2 percent to 4.7 percent of the total income of the lowest decile. Of course, the reductions would be smaller as we move up the income distribution, while for the higher-income classes, Pechman's estimates understate the tax rates if the transfer shares are overstated at these income levels. Thus, if the actual transfer shares by income classes are similar to those reported in Table 3, Pechman's 1980 results understate the progressivity of the tax system for the variant 1*c* incidence assumptions.

---

[6] I thank Gordon Green of the Bureau of the Census for making this unpublished data available to me.

[7] The Census Bureau reports that the lowest decile received $16.8 billion in in-kind transfers in 1980. Taking these into account, the bottom decile received 87 percent of its income in the form of cash and in-kind transfers. However, this figure is for the lowest decile of households as ranked by cash income only; the Census Bureau does not have estimates for rankings including cash and in-kind income. It should be mentioned that the Browning-Johnson and Pechman data are based on households ranked by income inclusive of all transfers.

[8] One other bit of evidence relevant to this issue is provided by the Treasury Department's *Blueprints for Basic Tax Reform* (1977). This study, undertaken to explore the possibilities of major tax reform using a consumption tax, developed a micro data base using a broad measure of family income for 1976. For the zero to $5000 economic income class, transfers were estimated to be 52 percent of economic income (Table 3). Since this income class contained more than 10 percent of households, the transfer share for the bottom decile would presumably be even higher.

To illustrate why the significance of the labor, capital, and transfer shares depends on incidence assumptions, consider sales and excise taxes. Pechman allocates these taxes in proportion to consumption outlays for each decile. Under this assumption, the sources of income are largely irrelevant (unless the marginal propensity to consume varies with the source of income), so changing the composition of income for each decile would not affect estimated tax burdens for these taxes. On the other hand, I have explained elsewhere (1985; Browning-Johnson) why I believe it is more appropriate to allocate sales and excise taxes in proportion to factor income. If that incidence assumption is used, changes in the composition of income imply changes in estimated sales and excise tax burdens. For example, a higher transfer share for the lowest decile then translates into a lower tax rate under sales and excise taxes. However, it is important to realize that the income composition error in Pechman's data has significant implications for estimated tax rates of other taxes even if this analysis of sales and excise taxes is not accepted. Tax rates for any taxes that are allocated to labor or capital income, or both, are affected when the shares of income from various sources are misestimated.[9]

---

[9] A careful comparison of the original figures with those provided by Pechman in his response to this note shows that the new overall rates do display substantially greater progressivity than the original ones, and the changes in the pattern of corporate, property, and payroll tax rates is pronounced. Although the *ad hoc* adjustments that have been made do change the results in the proper direction, there are two problems with the adjustments that lead me to believe the tax rates for the lower deciles are still overstated. First, the 1970 income shares are used as the basis for the adjustments. As I pointed out in fn. 4 above, the ratio of capital to labor income for the lowest decile is unacceptably high in the 1970 data. Since capital income is more heavily taxed than labor income, the result of overstating the capital share is to overstate the overall tax rate at the bottom. Note that, in his fn. 7, Pechman reports that the overall tax rate would be 2.5 percentage points lower for the bottom decile if the (more appropriate) 1966 shares are used. Second, the revised figures show the individual income tax rate for the bottom decile to have risen from 2.1 to 2.4 percent. Since the adjustments to the data increase transfers (which are not taxed under income taxes) and reduce factor income (which is partially taxed), it is difficult to see why the income tax rate did not fall. These two points suggest that the new estimates developed by Pechman have not fully taken account of the data problems I discussed.

REFERENCES

Browning, Edgar K., "Tax Incidence, Indirect Taxes, and Transfers," *National Tax Journal*, December 1985, *38*, 525–534.

______ **and Johnson, William R.**, *The Distribution of the Tax Burden*, Washington: American Enterprise Institute, 1979.

Musgrave, Richard A., Case, Karl E. and Leonard, Herman, "The Distribution of Fiscal Burdens and Benefits," *Public Finance Quarterly*, July 1974, *2*, 259–311.

Pechman, Joseph A., *Who Paid the Taxes, 1966–85?*, Washington: The Brookings Institution, 1985.

______ **and Okner, Benjamin A.**, *Who Bears the Tax Burden*, Washington: The Brookings Institution, 1974.

U.S. Department of the Treasury, *Blueprints for Basic Tax Reform*, Washington: USGPO, January 17, 1977.

Pechman's Tax Incidence Study: A Response

By JOSEPH A. PECHMAN*

I am indebted to Edgar Browning for calling attention to some peculiarities of the data underlying the estimates in my 1985 study. He is quite right that the ratios of transfer payments to income in the files for 1975 and later years are inconsistent with the corresponding ratios in the 1966 and 1970 files and in the *Consumer Population Surveys*, particularly in the lower part of the income distribution.

I do not believe, however, that the baby should be thrown out with the bathwater, as Browning seems to suggest. I made available to him detailed data on the tax burdens of labor income, capital income, and consumption by income classes which provided a basis for revising my calculations for 1975 and later years. Instead, he chose to draw inferences about the progressivity of U.S. taxes since 1966 on the basis of changes in the relative importance of the various taxes, which is not appropriate for this purpose. These inferences would be correct only if there were no changes in the structure of each tax and in the composition of income in the various income classes during the period studied (for example, if there were no changes in the progressivity of the individual income tax or in the distribution of dividends by income classes).[1]

It is straightforward to improve on this procedure. Where in the income distribution the share of transfer payments in total income is too low in 1975 and later years, the shares of labor and capital income are correspondingly high, and vice versa. To arrive at more realistic estimates of the shares for labor, capital, and transfer income, I first applied the 1970 shares to adjusted family income in each income decile in 1975, 1980, and 1985, and then made proportional adjustments in the columns and the rows alternately until they added to the correct totals.[2]

The incidence calculations for 1980 based on the revised weights are given in table 1 of my study, which shows the effective rates of the major taxes by deciles.[3] These calculations are based on the most progressive set of assumptions (variant 1*c*) presented in my study. Variant 1*c* assumes that the corporation income and property taxes are borne by capital in general, the payroll tax by labor, consumption taxes by consumers, and the individual income tax by those who pay it. I believe that this variant more nearly reflects the present state of incidence theory than any of the other variants.[4]

---

*The Brookings Institution, 1775 Massachusetts Avenue, N.W., Washington, D.C. 20036. I am grateful to Edgar Browning for helpful comments on an earlier version of this response, and to Charles E. Miles for programming the calculations.

[1] In fact, if this procedure were accurate, it would be necessary to prepare a MERGE file for only one year and then to project the incidence calculation for future years on the basis of changes in the relative importance of the various taxes used in the system.

[2] The income shares can be thought of as a matrix whose row entries add up to adjusted family income for each decile, and whose column entries add up to total capital, labor, and transfer income, respectively. As might be expected, when the 1970 shares are applied to later years, the totals do not add up to the known totals of capital, labor, and transfer income (the columns). These amounts were then adjusted proportionately to equal the known totals. This step yields incomes in each decile (the rows) which differed from the known totals, and this difference was eliminated by a proportional adjustment. These adjustments were repeated alternately until the columns (the totals of the income shares) and the rows (the total adjusted family income in each decile) added up to the known totals. For a proof that there is a unique solution to this procedure, see Michael Bacharach (1970, ch. 4). We have found that four or five iterations will yield satisfactory results, but we actually used as many as twenty to obtain the revised weights.

[3] I have not corrected the data base used in these calculations, because the costs would have been prohibitive. However, I believe that the adjustments of the weights for the various income shares described above provided a good approximation of the effective rates of the various taxes.

[4] Browning agrees, but he would go further. He would allocate consumption taxes to factor income rather than to consumption (see his paper, 1985), but he has not yet persuaded the profession of his position.

The revised calculations for 1980 give much more plausible estimates not only of the distribution of payroll taxes, but of other taxes as well. As in previous years, the effective payroll tax rate rises in the lower half of the income distribution, because transfer payments decline as a percentage of income; regressivity sets in thereafter. The U-shaped incidence by income classes of the corporation income and property taxes is also restored.[5]

Changes in the overall progressivity of the tax system for the period covered by the MERGE files are shown in table 2 of my study on the basis of the revised estimates. The distinct reduction from 1966 to 1985 in the effective tax rates in the top part of the distribution shown in the original calculations remains in the revised estimates. In 1966, those in the top decile paid 30.1 percent of their incomes in taxes; by 1985, their tax burden had declined to 26.4 percent. In the top percentile, the decline was from 39.6 percent to 25.3 percent. This was the result largely of the reductions in the corporation income tax and the decline in the importance of the property tax during this period.

On the other hand, the increase in tax burdens in the bottom part of the distribution shown in the original data disappears in the revised data. Although the effective tax rate in the bottom decile increases between 1966 and 1975, it declines between 1975 and 1985. On balance, the large increase in the payroll tax from 1966 to 1985 is just about offset by the effect of the (revised) increase in transfer payments in the first decile and more than offset in the second. However, if families were classified by market incomes rather than income plus transfer payments, the increase in the tax burdens at the lower end of the distribution resulting from the higher payroll taxes would be pronounced.[6]

Basically, the revised estimates show that there has been very little change in tax burdens throughout most of the income distribution, except at the very top where tax burdens have declined sharply because of reduced taxes on capital.[7]

A set of the revisions for the key tables for 1975, 1980, and 1985 is available on request.

---

[5] The relatively high income tax ratios in the lower deciles are the results of the accrual concept of income used in the analysis. Employee compensation includes employer contributions to pension plans in the year contributed, but the tax is deferred until pension benefits are received. Since there is no income to correspond to the employer contribution in the year the pension is received, the effective tax rate of persons receiving pensions is relatively high. For a discussion of this and other reasons why effective rates based on an annual accounting period may be distorted, see my study (pp. 48–50).

[6] The combined effect of transfer payments and taxation on the distribution of market incomes is shown in my study (table 4–7 and figure 4). On the basis of this classification, the tax transfer system is highly progressive, because transfers add to incomes in the bottom part of the distribution while taxes subtract from them in the top part. Unfortunately, we did not tabulate the 1966 and 1970 data by market incomes and therefore cannot make the comparison over the longer period 1966–85.

[7] Calculations using 1966 instead of 1970 income shares as weights raise progressivity for 1975, 1980, 1985, but the change from 1965–85 remains relatively small except at the very top of distribution. On the basis of the 1966 weights, the 1985 effective tax rates increase from 14.5 percent in the bottom decile to 24.1 percent in the sixth decile, and 26.0 percent in the ninth decile, and then decline to 25.6 percent in the top decile, and 25.2 percent in the top percentile.

REFERENCES

Bacharach, Michael, *Biproportional Matrices and Input-Output Change*, Cambridge: Cambridge University Press, 1970.

Browning, Edgar K., "Tax Incidence, Indirect Taxes and Transfers," *National Tax Journal*, December 1985, *38*, 525–534.

______, "Pechman's Tax Incidence Study: A Note on the Data," *American Economic Review*, December 1986, *76*, 1214–18.

Pechman, Joseph A., *Who Paid the Taxes, 1966–85?*, Washington: The Brookings Institution, 1985.

Race and Human Capital: Comment

By ROBERT A. MARGO*

According to human capital theory, changes in the racial schooling gap are a key factor in the historical evolution of black-white male income ratios. In a recent paper in this *Review* (1984), James Smith points out a basic paradox in the human capital explanation. Race differences in years of schooling have diminished sharply and continuously for male cohorts born in the twentieth century. Black-white male income ratios, however, rose only slightly in the aggregate before 1960. After 1960, the ratios increased appreciably. Smith resolves the paradox by constructing new estimates of the racial schooling gap for cohorts stretching back to the Civil War, based on retrospective educational attainment data from the 1940 and subsequent censuses. Race differences in years of schooling widened among males born from 1886 to 1910. Additionally, the quality of black schooling fell relative to the quality of white schooling. Since these cohorts dominate the census occupation and income statistics until 1960, Smith claims the relative constancy of black-white income ratios until 1960 is consistent with the human capital model.

This comment challenges one of Smith's conclusions. The increase in the racial schooling gap is shown to be spurious. Scholars often interpret census attainment data as a measure of years of schooling, but the data refer to highest grade completed. Historically, the average black pupil took longer than a single school year to complete a grade. Retention alone would not bias the census attainment data. Most blacks born in the late nineteenth century, however, were educated in ungraded schools. For them, census attainment data measure years of schooling, not grades. The shift from ungraded to graded schools took place throughout the period of educational retrogression identified by Smith. It is the change from years to grades that causes black schooling levels to appear to lag behind white schooling levels. Consistent data show a continuously decreasing racial schooling gap. Whatever the merit of Smith's explanation, it cannot rest on census attainment data.

Beginning in 1890, the U.S. Bureau of the Census reported school attendance rates for narrow age groups (for example, ages 5 to 9).[1] I use these data to construct new estimates of average years of schooling in the following manner. Let $p(j)$ = proportion of children of age j at school, $a(e)$ = minimum age at entering school, and $a(L)$ = maximum age at leaving school. I assume that $a(e)=5$ and $a(L)=20$, since attendance rates before age 5 or after age 20 were negligible for the period. Because data are unavailable for single years of age, I also assume the age-specific attendance rates are equal to the average attendance rate for the relevant age group.[2] The sum $\sum_{j=a(e)}^{j=a(L)} p(j)$ estimates average years of schooling.[3] Empirically, the number of students who skip a grade is less than the number who fail the grade. Hence average years of schooling should exceed average highest grade completed.

Table 1 presents the cohort-specific estimates of years of schooling. According to my calculations, the racial schooling gap fell from 3.8 years among 1886–90 cohorts to 2.5 years among 1906–10 cohorts. According

---

*Associate Professor of Economics, Colgate University, Hamilton, NY 13346, and National Bureau of Economic Research. I am grateful to Claudia Goldin, David Gray, and Paul Taubman for helpful comments.

[1] Attendance data are taken from U.S. Bureau of the Census (1943a, pp. 37–38).

[2] The bias arising from this assumption is very small. See my paper (1986b, fn. 7).

[3] I use attendance data for the census years 1890 to 1930. The $p(j)$ are linearly interpolated between census years. Estimates of years of schooling are constructed by following each cohort as it ages. For example, the 1891 cohort enters school in 1896 and leaves in 1911.

TABLE 1—THE BIAS IN EDUCATIONAL ATTAINMENT, MALE COHORTS, 1886–1910

Cohort	Years of Schooling	Highest Grade	Years per Grade
1886–90			
White	8.7	7.7	1.13
Black	4.9	4.7	1.05
Difference	3.8	3.0	0.08
1891–95			
White	9.0	8.2	1.10
Black	5.5	5.0	1.10
Difference	3.5	3.2	0.0
1896–1900			
White	9.5	8.7	1.09
Black	6.4	5.4	1.19
Difference	3.1	3.4	−0.10
1901–05			
White	9.9	9.2	1.08
Black	7.2	5.7	1.26
Difference	2.7	3.5	−0.16
1906–10			
White	10.5	9.7	1.08
Black	8.0	6.3	1.27
Difference	2.5	3.4	−0.17

Source: Years of schooling, see text; Highest grade: Smith (Table 4, p. 688); Years per grade = years of schooling/highest grade.
Note: Figures are cohort-specific averages.

to census attainment data, the racial schooling gap increased from 3 to 3.4 years. The two series flatly contradict each other.[4]

Which series is correct? Dividing my estimate of years of schooling by census attainment gives the number of years necessary to complete a grade. For white males, years per grade averages 1.1, with a slight downward trend. For black males, years per grade rises sharply over time. Among black males born from 1886 to 1890, it is 1.05, 0.08 years less than for whites. Among black males born from 1906 to 1910 it is 1.27, considerably higher than the white figure. According to Finis Welch (1973), black children in the early twentieth century took approximately two years to finish the first grade. Assuming they took somewhat more than a year per grade (but less than two years) to complete the remainder, 1.27 years per grade overall is not unreasonable. But the black figure for the 1886–90 cohorts—1.05 years per grade—is simply too low, not just absolutely, but relative to whites.[5]

Either my estimates of years of schooling for late nineteenth-century black cohorts are biased downward or the census attainment figures are biased upward. The available evidence suggests four reasons why census attainment data are at fault.[6] If education and mortality were negatively correlated, individuals surviving to 1940 would have higher average schooling levels than the average member of their birth cohort (Evelyn Kitagawa and Phillip Hauser, 1973; Sherwin Rosen and Paul Taubman, 1979). As schooling norms increase, individuals may overstate their true attainment, a bias known as educational creep (Edward Denison, 1962). My procedure is based on direct cohort-specific evidence; it is not affected by mortality or creep bias. Nevertheless, allowing for mortality or creep bias cannot fully explain the errors in the census schooling figures (see my 1986b paper).

Educational standards may have risen in black schools relative to white schools.[7] It seems odd, though, that relative standards would have improved at the same time the relative quality of black schools was getting worse. Furthermore, changing standards raises an index number problem. Which set of standards do we use for racial comparisons? My calculations show the choice of a standard—specifically, years per grade—does not matter for whites. The choice matters a great deal for blacks. But as long as a constant standard is used to transform years to grades for blacks, the racial schooling gap decreases over the period. For example, if a middle-of-period standard (1.19 years per grade) is used, the racial schooling gap is 3.6 years $(7.7-(4.9/1.19))$ for the 1886–90 cohorts and 3.0 years $(9.7-(8/1.19))$ for the 1906–10 cohorts.[8]

---

[4] Results for females are similar, and are available from the author.

[5] Data on age-specific months of school attendance in 1890 and 1900 show that if census attainment data are to be believed, black males born from 1886 to 1890 could complete a grade in 31 percent fewer months than white males (see my 1986b paper). The apparent relative efficiency of black students is inconsistent with the poorer quality of black schools and with socioeconomic factors such as poverty and adult illiteracy that severely hindered black school performance (see my 1986a paper).

[6] Black school attendance may have been underreported in 1890 and 1900, although the Census Bureau believed the bias was small (see U.S. Bureau of the Census, 1918, p. 378). Reasonable adjustments for bias (for example, estimating attendance rates in 1900 by extrapolating backwards at the rate of growth prevailing from 1910 to 1920) do not change the substantive results. Details are available from the author on request.

[7] The black attendance rate ages 5 to 20 increased by 29 percentage points from 1900 to 1930. The enrollment of less able students may have caused years per grade for blacks to rise. The problem with this argument is the white attendance rate also rose from 1900 to 1930 (by 18 percentage points), yet years per grade for whites fell slightly (see Table 1).

There is, however, a much simpler explanation for the bias. The Census Bureau asked for highest grade completed, but realized that many individuals had attended ungraded schools. For such people, enumerators recorded the "number of years the person attended school."[9] Comprehensive national data are unavailable, but there is no doubt that ungraded schools were more common in the South, especially in black school districts. In Texas, for example, 89 percent of black schools in 1900 were ungraded.[10] The transition from ungraded to graded schools took place fairly rapidly after the turn of the century, and although an exact date cannot be determined, seems to have been completed by World War I. Blacks entering school after 1910 were educated in graded schools, but the previous generation was not. Adjusting for changes in definition shows that, whether measured by years or grades, the racial schooling gap must have fallen among cohorts born from 1886 to 1910.

This comment has shown the divergence in racial schooling levels emphasized by Smith is a consequence of biases in census attainment data. Decreases in the relative quality of black schooling, as Smith and others have pointed out, were deep and prolonged.[11] Whether these changes in relative quality are sufficient to reverse my findings remains to be seen.

---

[8] Using a constant standard understates the reduction in the racial schooling gap over time since it does not adjust for increases in the length of the school year in black schools relative to the national average (see my 1986b paper).

[9] For example, a pupil attending school from age 6 to age 12 would give his attainment as 6 years, which would be recorded as the sixth grade. To avoid overstating schooling levels among individuals who attended ungraded schools, enumerators were supposed to transform years to grades whenever possible, but my calculations suggest few did so in practice. See U.S. Bureau of the Census (1943b, p. 178).

[10] Calculated from data in Superintendent of Public Instruction, State of Texas (1900, pp. 71–85).

[11] See my 1985 book for quantitative evidence on race differences in school quality at the turn of the century.

REFERENCES

Denison, Edward, *The Sources of Economic Growth in the United States and the Alternatives Before Us*, New York: Committee for Economic Development, 1962.

Kitagawa, Evelyn M. and Hauser, Phillip H., *Differential Mortality in the United States*, Cambridge: Harvard University Press, 1973.

Margo, Robert A., *Disfranchisement, School Finance, and the Economics of Segregated Schools in the United States South, 1890–1910*, New York: Garland Press, 1985.

______, (1986a) "Educational Achievement in Segregated School Systems: The Effects of 'Separate-but-Equal'," *American Economic Review*, September 1986, *76*, 794–801.

______, (1986b) "Race, Educational Attainment, and the 1940 Census," *Journal of Economic History*, March 1986, *46*, 189–98.

Rosen, Sherwin and Taubman, Paul, "Changes in the Impact of Education and Income on Mortality in the United States," in Linda Del Bene and Fritz Scheuren, eds., *Statistical Uses of Administrative Records With Emphasis on Mortality and Disability Research*, Washington: USGPO, 1979.

Smith, James P., "Race and Human Capital," *American Economic Review*, September 1984, *74*, 685–98.

Welch, Finis, "Education and Racial Discrimination," in Orley Ashenfelter and Albert Rees, eds., *Discrimination in Labor Markets*, Princeton: Princeton University Press, 1973.

Superintendent of Public Instruction, State of Texas, *Report*, Austin: State Printers, 1900.

U.S. Bureau of the Census, *Negro Population in the United States, 1790–1915*, Washington: USGPO, 1918.

______, (1943a) *Sixteenth Census of the United States: 1940, Population*, Vol. II: *Characteristics of the Population, Part 1: United States Summary*, Washington: USGPO, 1943.

______, (1943b) *Sixteenth Census of the United States: 1940, Population*, Vol. IV: *Characteristics by Age, Part 1: United States Summary*. Washington: USGPO, 1943.

Race and Human Capital: Reply

By James P. Smith*

To arrive at a tenable interpretation of black history, it is important to resolve the conflict between my measure—retrospective years of schooling completed—and Robert Margo's measure—prospective accumulation across ages of cohort-specific attendance rates. My series reconciled the apparent inconsistency between the stagnant pre-1940 black-white income ratios and what scholars had previously thought was a steady narrowing of racial education differences. Margo's series not only would restore that inconsistency, but also would rewrite a little history itself, since it implies that the first two decades of this century witnessed an unprecedented boom in black education. If Margo is right, economists should hesitate to use (for racial comparisons) the education data in the recently released 1940 and 1950 census micro files. If I am right, historians, forced by data availability to rely heavily on school attendance rates to track educational advances, may have seriously misrepresented between-group comparisons. Since our two series could hardly be more different, and both are free from computational errors, which is right?

In this reply, I make three basic points: 1) Reporting biases in census school attendance data distort the true school-accomplishment profiles of cohorts; 2) school attendance rates are inherently biased measures of racial differences in school completion; and 3) the reasons for blacks' less rapid schooling advances are largely independent of whether southern black schools changed from an ungraded to a graded system during these years.

The Census historical series on school attendance contains two reporting biases. The relatively minor one concerns the truncation at age 20, which systematically understates the true growth in white male education. In 1930, 16 percent of 20-year-old white men attended school during the previous year. Among the birth cohorts under dispute here, the proportion of white men who attended college almost doubled, from 9 to 17 percent. Since his attendance series stops at age 20, Margo ignores most of the white male gains at the college level.

The more damaging reporting bias involves a temporary switch in the census school attendance question. At all censuses prior to 1890 and after 1900, the question was whether the respondent attended school during the previous year. In 1890 and 1900, however, respondents were only asked the number of months they attended school. Those who reported a month were coded by the census as school attendees. The serious comparability issues that resulted were aptly summarized by the Bureau of the Census:

> It is undoubtedly true that in many cases the fact that a certain person has attended school during the preceding years may be known, while the number of months of his attendance at school may be unknown, especially when the person interviewed is not the one to whom the information apply. In such cases, enumerators would often be disposed to leave the answer blank, rather than to enter the fact of attendance stating that its duration was unknown. The returns for 1910 are therefore doubtless more complete than those of 1900 and the apparent increase in the proportion attending school for 1910, as compared with 1900, as given in Table 12, is due in part—probably in large part—to the greater accuracy of the returns. [1913, p. 1104]

It is very likely that the largely illiterate, rural, southern black parents found it difficult to answer this revised census atten-

---

*Senior Economist, The Rand Corporation, 1700 Main Street, Santa Monica, CA 90406. Financial support from the National Science Foundation and the research assistance of Judith Fischer are gratefully acknowledged.

dance question accurately. The consequences of the revision are apparent in the time-series trends. After black school attendance rates jumped sharply—from 9.2 percent in 1870 to 33.3 percent in 1880—they actually dropped to 31.1 percent and 29.6 percent in the next two censuses with the revised question. With the return of the original question in 1910, black attendance rates jumped sharply to 43.6 percent and rose to 53 percent in 1920.

The 1890 and 1900 revised question clearly resulted in a substantial understatement of school attendance during those years, a bias that affects the birth cohorts at issue unequally. For example, Margo's 1886–90 years of schooling are almost completely derived from the underreported 1880 and 1890 census numbers, while his 1906–10 birth cohort's schooling did not use these two censuses at all. Because the very substantive issue under debate concerns irregularity of the timing of black schooling advances during these decades, the magnitude of the bias in Margo's series is unknowable. However, if we take as an extreme the census conjecture that the increase in school attendance for both races was due to misreporting, 60 percent of Margo's reported 1.3 rise in black schooling could be due to census reporting biases in the school attendance question.[1]

Although census reporting bias is important, adjusting for it does not fully reconcile the two series. The second problem is that Margo needs a great deal more information to correctly map from attendance rates to school completion. Although he labels his series "years of schooling," it literally is the average number of years that cohorts were enrolled in school for at least a day. To translate into years of schooling completed, one must also know enrollment by grade as well as grade retention rates. In his comment, Margo correctly states that grade retention alone would not bias my census attainment data. But grade retention does bias his attendance-based series by overstating black schooling advances relative to whites. If γ is the average number of enrolled years it takes to complete a grade, education completed would be $\sum_{j=a(e)}^{j=a(L)} p(j)/\gamma$.-s) Margo's secular change in schooling should be scaled down by the race-specific γ's. As long as there existed race differences in grade retention rates, a fact that nobody disputes, even racially neutral advances in school attendance rates would result in larger white male gains in education completion. Consequently, the more rapid black improvements in school attendance, reported by Margo, are not necessarily inconsistent with more rapid white improvements in grade completed.

We simply do not know enough about historical trends in grade retention to map school attendance into grade completion. Finis Welch's (1976) rough estimates imply average grade retention rates of 1.54 for blacks and 1.21 for native whites, a race difference that does not seem unreasonable.[2] Applying these retention rates to Margo's series would by itself reduce his estimate of the black gain over whites to 0.6 of a year. If we then add in the impact of reporting bias, Margo's relative black gain would be less than 0.2 of a year (instead of his 1.3 years reported).

This bias is present even if retention rates were constant across cohorts. But Welch (1973) also presents evidence that black retention rates rose over this period. For first graders, retention rates among those born in 1888 were 1.37; by the birth year 1909 they rose to 1.96. This spectacular jump in reten-

---

[1] Margo would then have calculated, for the 1886–90 cohort, 6.81 (black) and 10.03 (white) years of schooling (a difference of 3.2 instead of his reported 3.7). His new numbers for the 1906–10 birth cohort are 10.81 and 8.8, so that the relative black gain across these cohorts would be 0.49 years instead of his reported 1.3 years. These revised numbers also include a small adjustment for the age 20 truncation. Contrary to Margo's contention, this bias has a differential racial impact even when the reporting bias affects both races. The example in the text assumes that the increase in school attendance for both races between 1890 and 1910 was due to reporting bias. To the extent that blacks were differentially affected, a larger proportion of Margo's reported rise in black schooling is a consequence of reporting bias.

[2] I evaluated Welch's numbers at the mean grade completion levels for the 1886–90 birth cohort in Table 1.

TABLE 1—MEAN SCHOOLING LEVELS BY BIRTH COHORT

Birth Cohort	All Whites	Blacks	Native-Born Whites	Foreign-Born Whites	Born in South Whites	Born in South Blacks	Born in North Whites	Born in North Blacks
1951–55	13.31	12.81	13.37	12.23	12.90	11.94	13.56	12.54
1946–50	13.37	11.75	13.41	12.45	12.77	11.50	13.69	12.24
1941–45	13.34	11.47	13.40	12.29	12.63	11.19	13.75	12.15
1936–40	12.66	10.68	12.72	11.56	11.91	10.36	13.11	11.71
1931–35	12.52	10.17	12.58	11.43	11.62	9.79	13.05	11.50
1926–30	12.05	9.24	12.08	11.44	11.02	8.75	12.55	11.94
1921–25	11.97	8.77	12.01	11.33	10.89	8.36	12.49	10.76
1916–20	11.36	7.80	11.37	10.99	10.22	7.37	11.84	10.41
1911–15	10.90	7.16	10.92	10.58	9.75	6.78	11.41	9.84
1906–10	10.46	6.58	10.55	9.78	9.29	5.21	11.09	9.29
1901–05	9.46	5.80	9.60	8.57	8.56	5.46	10.07	8.21
1896–1900	8.97	5.56	9.27	7.58	8.27	5.20	9.68	8.02
1891–95	8.40	5.00	8.86	6.72	7.86	4.68	9.27	7.64
1886–90	7.89	4.65	8.45	6.28	7.45	4.33	8.86	7.26
1881–85	7.62	4.33	8.12	6.25	7.10	3.96	8.53	7.23
1876–80	7.40	4.01	7.83	6.16	6.83	3.77	8.26	6.51
1871–75	7.24	3.72	7.61	6.07	6.48	3.55	8.08	5.79
1866–70	7.10	3.39	7.43	6.23	6.45	3.26	7.78	4.77
Pre-1865	6.81	2.62	7.11	6.06	6.02	2.42	7.49	4.52

tion is not surprising, as it coincides in timing with the reduction in the quality of black southern schools. Less than a 5 percent secular increase in black grade retention rates (well below these estimates) would completely eliminate the remaining relative black gain in the Margo series. My point is not that reasonable retention numbers exist that would map Margo's series into mine, although that is certainly true. Rather, without knowing these retention parameters, creating a grade completion series from school attendance rates can be treacherous.

In his comment, Margo offers one substantive argument on why my series might be biased: the change from an ungraded to a graded system in the South. While one might want more documentation on timing and magnitude than reference to a single southern state, Margo's emphasis on events in the South misrepresents the historical processes at work. To see this, Table 1 presents revised estimates of average schooling completed, by five-year birth cohorts.[3] To clarify the forces at play, I also list mean schooling levels for native-born and foreign-born whites and by region of birth for both races. To isolate differential secular trends, Table 2 lists race differences in school completion for each birth cohort.

One major historical force was foreign immigration. During the first fifteen years of this century, 13 million immigrants came to the United States, but then foreign immigration dropped precipitously. The calendar years 1901–15 intersected with the peak 15–24 ages of immigration for the 1886–90 cohort, but our 1906–10 birth cohort completely missed the heavy immigration years. As a result, the proportion of foreign-born whites dropped sharply across these birth cohorts, from 25 to 10 percent (see the third column of Table 2). Margo's prospective series, unfortunately, leaves out most of the foreign born.[4] The uneven impact of im-

---

[3] My original series used published census data for the 1940 and 1950 censuses, the 1979 *Current Population Series*, and the 1960 and 1970 census micro files. My revised series substitutes the recently released micro data files for the 1940, 1950, and 1980 censuses. In all other respects the construction of my revised series is identical to that described in my earlier paper.

[4] To illustrate, among the 1886–1900 cohort, the percentage of white men who were foreign born were

TABLE 2—RACIAL DIFFERENCES IN MEAN SCHOOLING LEVELS
(Additional years of schooling of whites)

	All	Native Born	Percent Native-Born Whites	Born in South	Born in North	Percent Blacks Southern Born
1956–60	0.71	0.77	95.2	0.57	0.72	59.1
1951–55	1.13	1.23	94.8	0.95	1.02	64.4
1946–50	1.61	1.68	95.1	1.27	1.45	67.8
1941–45	1.86	1.98	94.6	1.44	1.60	73.9
1936–40	1.98	2.09	94.6	1.55	1.40	77.9
1931–35	2.35	2.45	94.6	1.83	1.55	78.2
1926–30	2.80	2.85	95.5	2.27	1.51	79.1
1921–25	3.20	3.26	94.9	2.53	1.73	82.7
1916–20	3.55	3.61	95.8	2.85	1.43	87.4
1911–15	3.74	3.80	93.7	2.96	1.57	88.9
1906–10	3.88	4.01	90.4	3.08	1.80	89.3
1901–05	3.67	3.86	86.2	3.10	1.86	89.4
1886–1900	3.41	3.78	82.6	3.07	1.66	89.9
1891–95	3.40	3.92	78.4	3.18	1.63	90.7
1886–90	3.25	3.86	74.4	3.13	1.60	91.6
1881–85	3.29	3.90	73.5	3.14	1.31	91.4
1876–80	3.40	3.87	74.7	3.06	1.75	92.0
1871–75	3.52	3.92	76.0	2.94	2.29	93.1
1866–70	3.71	4.06	72.7	3.19	3.00	93.0
Pre-1865	4.18	4.49	–	3.60	2.97	–

migration on these birth cohorts can be seen by comparing the native born and all white columns in Tables 1 and 2. Across our disputed birth cohorts, native whites gained only 0.15 of a year of schooling more than blacks, in contrast to the 0.6 of a year net gain for all whites.

Three-quarters of my relative white education gain, therefore, reflected foreign immigration. Immigration accelerated the across-cohort rise in white male education, because the weight assigned to the more highly educated native-born whites increased rapidly during these years. Table 1 also demonstrates that the education of foreign-born whites expanded sharply, substantially narrowing their disparity with native-born whites. The better education opportunities in America imparted clear schooling benefits for the foreign born. Only 18 percent of immigrants in the 1886–90 birth cohort were domestically educated, compared with almost one-half the immigrants in the 1906–10 birth cohort.[5]

Events in the North, and not the South, drive my remaining white education gains. Table 2 indicates that racial differences were basically stable for those born in the South, but expanded slightly for the northern born during the retrogression era. The key historical event was an acceleration of schooling in the North, not an attenuation of black schooling in the South.

In the North, the principal development was the spread of the public high school on a wide scale. By the 1916 birth cohort, more than 8 in 10 white male elementary graduates continued in school and 60 percent would earn their high school diplomas—almost double the rates that existed 30 years earlier

---

4.5 (ages 10–14), 17.9 (ages 20–24), 22.5 (ages 30–34). By the 1906–10 cohort, the corresponding percentages were 3.5, 7.8, and 7.7. Margo's series includes only the youngest immigrants. The complications introduced by immigration are a major disadvantage of Margo's prospective measure, compared with my retrospective education. Because his series is dominated by the native born, his series would have been almost identical if he had limited it to native-born whites.

[5] This is a rough estimate, calculated by taking the ratio of numbers of foreign born at ages 10–14 to those foreign born at ages 40–44.

(see my earlier paper). If we gave black men the whites' region-specific increases in education, during the retrogression era, white male schooling would still have risen a quarter of a year more than black schooling did. Consequently, my relative white education gains can be fully explained by two unique historic events: immigration and the expansion of the northern high school. Neither event was affected by any transition from ungraded to graded schools in the South.

My emphasis should not be misconstrued as an argument that political events in the South had no impact on black schooling. Tables 1 and 2 do indicate that these were still difficult years for black education in the South. Black-white education differences were stable during the retrogression era, in sharp contrast to the substantial education advances southern blacks would subsequently enjoy. Indeed, I would interpret Margo's expanding black attendance rates relative to my more stable racial differences in attainment as an additional barometer of the deteriorating quality of black schooling. If one consequence of the deteriorating quality was increasing black grade retention, one must find rising relative black attendance coexisting with stable race differences in grade completed.

REFERENCES

Margo, Robert A., "Race and Human Capital: Comment," *American Economic Review*, December 1986, *76*, 1221–24.

Smith, James P., "Race and Human Capital," *American Economic Review*, September 1984, *74*, 865–98.

Welch, Finis, "Ability Tests and Measures of Differences Between Black and White Americans," WD–2102–HEW, Rand Corporation, October 1976.

______, "Education and Racial Discrimination," in Orley Ashenfelter and Albert Rees, eds., *Discrimination in Labor Markets*, Princeton: Princeton University Press, 1973.

U.S. Bureau of the Census, *Thirteenth Census of the United States*, Vol. 1, *Population, 1910*; *General Report and Analysis*, Washington: USGPO, 1913.

NOTES

The annual meeting of the American Economic Association will be held in New Orleans, Louisiana, December 28–30, 1986.

The Professional Placement Service will be located at the Sheraton New Orleans Hotel. It will be open from 10:00 A.M. to 5:00 P.M., December 27; 9:00 A.M. to 5:00 P.M., December 28–29; and 9:00 A.M. to 12:00 noon, December 30.

Call for Papers for the 1987 *AEA Meetings*

Members wishing to give papers or make suggestions for the program for the meetings to be held in Chicago, IL, December 28–30, 1987, are invited to send their ideas to Robert Eisner, Department of Economics, Northwestern University, Evanston, IL 60201. As in the past, the sessions sponsored by the American Economic Association will be of two types: invited papers and contributed papers. Most of the invited papers will be published in the *Papers and Proceedings* issue to appear May 1988, the contributed papers will not be so published. Suggestions of topics or proposals for papers for the invited sessions should be submitted as soon as possible. Abstracts of proposed contributed papers must be received no later than February 1, 1987. Econometric studies or highly mathematical papers are not appropriate for the sessions sponsored by the American Economic Association. However, members wishing to present such papers may submit their proposals or abstracts to the Econometric Society which meets with the American Economic Association, and normally schedules a number of sessions involving such papers.

The Luxembourg Income Study (LIS) has gathered in one central location (The Center for Population, Poverty and Policy Studies in Luxembourg), and made comparable, several recent large microdata sets that contain comprehensive measures of income and economic well-being for a set of modern industrialized welfare states. Because of the breadth and flexibility afforded by the microdata, researchers are free to make several choices of perspective (definition of unit: family, household, etc.; measure of income: earnings, after tax and transfer income, etc.; and population to be studied: males, females, urban families, elderly households, etc.) within the same research paper. This microdata set creates a potentially rich resource for comparative applied research in economics and public policy. The LIS databank currently covers Australia, Britain, Canada, Germany, Israel, Norway, Sweden, and the United States, with France, Netherlands, and Switzerland soon to be added. A copy of the User Guide and further information can be obtained by writing either Professor Tim Smeeding, Economics and DSSR; 2120 Annex; University of Utah; Salt Lake City, UT 84112, or Professor Lee Rainwater, Sociology; 530 Wm. James Hall; Harvard, University; Cambridge, MA 02138.

The 1987 annual meeting of the Eastern Finance Association will be held April 22–25, in Baltimore, Maryland. Address program inquiries to Gershon N. Mandelker, Graduate School of Business, Mervis Hall, University of Pittsburgh, Pittsburgh, PA 15260, and local arrangement inquiries to Robert Schweitzer, College of Business and Economics, Department of Business Administration, Newark, DE 19716.

The Center for Education and Research in Free Enterprise at Texas A&M University and the Walter Eucken Institute, Freiburg, Germany, will jointly sponsor a program, "The Hayek Symposium on Knowledge, Information, and Competition," to be held June 20–23, 1987, in Freiburg. Participation will be by invitation only and limited to twenty scholars. Preference for presenting papers will be given to research in progress by assistant and associate professors. Selected papers will be published in *The Hayek Symposium Series.* All travel and hotel expenses will be paid by the cosponsors. For more information, write to Dr. Steven Pejovich, The Center for Free Enterprise, 459 Blocker, Texas A&M University, College Station, TX 77843, or The Hayek Symposium, Walter Eucken Institute, Goethestrasse 10, Freiburg Im Breisgau, West Germany.

Call for Papers: The research annual, *Advances in Econometrics* (JAI Press), is soliciting papers in the areas of Robust and Nonparametric Statistical Inference, to be published in 1988. Send abstracts no later than March 1, 1987, to either editor: Professor George F. Rhodes, Jr., Department of Economics, Colorado State University, Fort Collins, CO 80523; Professor Thomas B. Fomby, Department of Economics, Southern Methodist University, Dallas, TX 75275.

Call for Papers: The Conference on U.S.–Asia Economic Relations, sponsored by the Committee on Asian Economic Studies (CAES) will be held June 21–24, 1987, in Hong Kong, in cooperation with the Chinese University of Hong Kong. Papers on Asian economic studies, especially intercountry research and U.S.–Asia economic interaction are invited. Abstracts (200 words) must be received no later than January 31, 1987. Invited conferees will be provided accommodations for four nights and food for conference days only. Contact Professor M. Jan Dutta, Director, CAES, Department of Economics, Rutgers University, New Brunswick, NJ 08903.

Call for Papers: Siena College is sponsoring a multi-disciplinary conference on the 50th anniversary of World War II. The focus is 1937—however, papers dealing with broad issues of earlier years are welcome. Contact Thomas O. Kelly II, Head, Department of History, Siena College, Loudonville, NY 12211.

Call for Papers: The University of Florida is sponsoring a new student-run and faculty-edited journal: *Journal of Law and Public Policy*. For full information, or to submit papers, please write to the Editors, *JL&PP*, College of Business Administration, University of Florida, Gainesville, FL 32601.

Call for Papers: The first National Conference of the Association for the Study of Food and Society (ASFS) will be held April 1–3, 1987, at Michigan State University in East Lansing. The theme is "Solving Global Food Problems: A New Beginning." Scholars in all disciplines are invited to send abstracts (or completed papers) no later than January 1, 1987, to Professor Bill Whit, President, ASFS, Department of Sociology, Aquinas College, Grand Rapids, MI 49506 (telephone 616+459–8281, ext. 217).

Call for Papers: The International Society for Inventory Research will hold a conference on the Economics of Inventory Management at Wesleyan University, June 16–18, 1987. For further information, contact Michael C. Lovell, Department of Economics, Wesleyan University, Middletown, CT 06457.

The National Association of Forensic Economists will publish a journal, *The Journal of Forensic Economics*, twice a year. For more information, write the National Association of Forensic Economists, P.O. Box 30067, Kansas City, MO 64112.

The Census Bureau's third annual Research Conference will be held March 29–April 1, 1987, in Baltimore, Maryland. The theme is "Improving the Quality of Survey Results." For full information, contact Carolee Bush, Office of the Director, Bureau of the Census, Washington, D.C. 20233 (telephone 301+763–4900).

The annual Congress of the International Institute of Public Finance, "Public Finance and Performances of Enterprises," will be held August 24–28, 1987, in Paris, France, and will be their semicentennial. The Chairman of the Program Committee is Professor Manfred Neumann, University of Erlangen-Nürnberg, Lange Gasse 20, D 8500 Nürnberg, West Germany. For further information, contact IIPF President: Karl Roskamp, IIPF, Saar University, D 6600 Saarbrücken 11, West Germany.

The Society for Historians of the Early American Republic will hold its annual meeting, July 16–18, 1987, at the Temple University City Center in Philadephia. The Program Chair is Dr. James B. Stewart, Macalester College, St. Paul, MN 55105; the Secretary-Treasurer is James H. Broussard, Lebanon Valley College, Annville, PA 17003.

The American Antiquarian Society (AAS) offers short- and long-term Visiting Research Fellowships for June 1, 1987 to May 31, 1988. The National Endowment for the Humanities (NEH) Fellowships are for six to twelve months, with a maximum stipend of $27,500. The NEH Fellowships may not be awarded to degree candidates or for study leading to advanced degrees. NEH Fellows must devote full time and may not hold any other major fellowships, except sabbaticals or other grants from their own institutions. Short-term Peterson Fellowships are from one to three months, with a maximum stipend of $700 per month. They are open to individuals at work on doctoral dissertations, who are engaged in research and writing on any field in American history and culture through 1876. For the same stipend are two Frances Hiatt Fellowships for one or two months' graduate work at the AAS. Also offered jointly with the Newberry Library in Chicago are short-term fellowship programs. The deadline for receipt of applications is January 31, 1987. Full information and forms should be requested from John B. Hench, Associate Director for Research and Publication, AAS, 185 Salisbury Street, Worcester, MA 01609 (telephone 617+752–5813 or 617+755–5221).

The Center for International Affairs, Harvard University, announces the Harvard Academy for International and Area Studies. The Academy consists of senior and junior scholars dedicated to training in area studies. The Academy's junior scholars will be known as Ira Kukin Scholars. The senior scholars will act as mentors. Applications for the 1987–88 class of Ira Kukin Scholars are due by February 13, 1987, and must include: a copy of each graduate transcript; a statement not to exceed 1,000 words of the applicant's research and intellectual objectives over the next two years; a statement of how this particular appointment will contribute to the applicant's career goals, not to exceed 500 words; a current curriculum vitae, including biography;

and three letters of recommendation. Send all materials to Dr. Chester Haskell, Executive Officer, Center for International Affairs, 416B Coolidge Hall, 1737 Cambridge Street, Cambridge, MA 02138 (telephone 617+495-2965).

Harvard University's Center for International Affairs and Center for European Studies, in collaboration with the Ford Foundation, announces dissertation and postdoctoral support for research on the relationship between European society and Western security. This program places special emphasis on major policy issues and alternatives facing Europeans in the field of security, broadly defined, and the internal factors that influence European choices among these alternatives. The centers are particularly interested in building bridges among the social sciences and between European studies and strategic/defense studies. Fellows selected will spend the 1987–88 academic year at the Centers. The Dissertation Fellowship carries a stipend of $11,000 and the Postdoctoral Fellowship a stipend of $22,000. The deadline for applications is February 17, 1987. For more information, contact Fellowship Office, Room 402, Center for International Affairs, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138 (telephone 617+495-1669).

The Thornton D. Hooper Fellowship Competition seeks to promote original research on issues affecting U.S. and international security matters. The selected applicant will spend a year in residence at the Foreign Policy Research Institute, beginning fall 1987, and will receive a stipend commensurate with his or her experience. Applicants should submit copies of three publications (or writing accepted for publication), a curriculum vitae, three letters of recommendation, and a 6-page, double spaced research proposal. Fellows must have received or be about to receive a doctoral degree, or have equivalent experience. The deadline is February 16, 1987. Submit applications to Hooper Fellowship Competition, Foreign Policy Research Institute, 3508 Market Street, Suite 350, Philadelphia, PA 19104.

The P.W.S. Andrews Memorial Prize of £300 is awarded annually for an essay within the general field of Industrial Economics and Theory of the Firm, broadly interpreted. Entry is open to those under age 30 (or within 8 years of taking a first degree). The essay, written in English, and not exceeding 10,000 words, should be original research by the candidate only, not previously published, and not previously awarded any other prize. The winning essay may be published in the *Journal of Industrial Economics*. The deadline is December 31. Details are available from Miss B. Cox, University House, University of Lancaster, Bailrigg, Lancaster, LA1 4YW, England, or the General Editor, *JIE*, Institute of Economics and Statistics, University of Oxford, St. Cross Bldg, Manor Road, Oxford, OX1 3UL, England, or the North American Editor, New York University, Graduate School of Business, 90 Trinity Place, New York, NY 10006.

Pre- or postdoctoral graduate students are invited to submit papers (no longer than 35 pages, double spaced) for the 1987 Dorothy S. Thomas Award competition, established by the Population Association, in the fields of internal migration or the interrelationships among social, economic, and demographic variables. The award is $1,500. Five copies of the paper (and nominating letters sent directly from the references) should be sent to Julie DaVanzo, The Rand Corporation, 1700 Main Street, Santa Monica, CA 90406, and must arrive no later than January 15, 1987.

The annual dissertation competition sponsored by the American Real Estate and Urban Economics Association in cooperation with the Homer Hoyt Advance Studies Institute will award one or more doctoral dissertation prizes of $2,000 in the field of real estate and applied urban or land economics. For full particulars, contact the chairman of the Dissertation Review Committee: Professor John B. Corgel, Department of Real Estate, College of Business, Georgia State University, Atlanta, GA 30303.

A conference, "Private Sector Involvement in Health Care: Implications for Access, Cost, and Quality," will be held at SRI International, April 9–10, 1987, at Menlo Park, California. The conference is supported by Project Hope, University of California-Berkeley, Medical College of Virginia, and SRI International. For further information, contact Nelda McCall, Director, Health Policy Research, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025 (telephone 415+859-3679).

The 1987 annual meeting of the Urban Affairs Association will be held in Akron, Ohio, April 22–25. The theme is Cities in Transition. For further information, contact Mary Helen Callahan, Executive Director, UAA, University of Delaware, Newark, DE 19716 (telephone 302+451-2394).

The editors of *Social Philosophy & Policy* announce that Volume 6, issues 1 ("Capitalism") and 2 ("Socialism"), are open to unsolicited manuscripts. For full information, write to John Ahrens, Managing Editor, *SP&P*, Social Philosophy and Policy Center, Bowling Green State University, Bowling Green, OH 43403.

The Center for Russian and East European studies, University of Michigan, announces a grant from the Andrew W. Mellon Foundation to offer one or more postdoctoral fellowships in any area of Russian and East European studies. The stipend is $18,000 and may be used in lieu of salary or in conjunction with other fellowship awards or research grants. Fellows must be in residence in Ann Arbor for the major portion of their fellowship terms, and participate in Center activities. The deadline for applications is February 15, 1987. Please submit a full curriculum vitae, a statement of current research and teaching interests, and a proposal describing how these interests would be pursued under an award. Submit these and three letters of recommendation directly to Roman Szporluk, Director, Center for Russian and East European Studies, University of Michigan, 208 Lane Hall, Ann Arbor, MI 48109.

The Institute of American Cultures at UCLA offers fellowships to support study of Afro-Americans, American Indians, Asian Americans, or Chicanos. The stipend for graduate fellowships is $10,000 per year plus registration fees and out-of-state tuition, if applicable. This fellowship is available to applicants who have applied to or are enrolled at UCLA. Postdoctoral fellowship stipends are $20,000–$25,000 per year. For further information and applications, contact the Fellowship Director of the appropriate ethnic center at UCLA, Campbell Hall, Los Angeles, CA 90024: Center for Afro-American Studies, RM 3111; American Indian Studies Center, RM 3220; Asian American Studies Center, RM 3232; Chicano Studies Research Center, RM 3121.

Call for Papers: The *Social Science Quarterly* solicits manuscripts in economics which are directed to professional economists, and yet of interest to other social scientists. Brief research notes are encouraged, and papers longer than 30 pages are not acceptable. Submit 4 copies to Charles M. Bonjean, Editor, *SCQ*, W. C. Hogg Bldg, University of Texas, Austin, TX 78713.

New Journal — Call for Papers: The Korea International Economic Association announces publication of an academic, general purpose, refereed periodical: *International Economic Journal*. Send manuscripts to Professor Wontack Hong, Managing Editor, *IEJ*, Department of International Economics, Seoul University, Seoul 151, Korea, or to Professor Young Chin Kim, Department of Economics, Northern Illinois University, DeKalb, IL 60115.

Call for Papers: *Studies in Economic Analysis* is a biannual, refereed, student-edited journal soliciting research articles. The submission fee is $4.00 for nonsubscribers. Send manuscripts to the Editors, *SEA*, Department of Economics, College of Business Administration, University of South Carolina, Columbia, SC 29208.

Call for Papers: A Risk Theory Seminar will be held April 23–25, 1987, at the University of Texas, Austin. Persons interested in participating should send a detailed outline of their proposed presentation to Professor Emilio Venezian (RTS Secretary), Department of Business Administration, Rutgers University, Bradley Hall, Newark, NJ 07102, or Professor S. Travis Pritchett (RTS Chairperson) College of Business Administration, University of South Carolina, Columbia, SC 29208. The deadline is December 15, 1986.

Call for Papers: The Association for Canadian Studies in the United States (ACSUS) will hold its biennial meeting at the Bonaventure Hotel in Montreal, October 8–10, 1987. Panels will be organized on all aspects of Canadian studies and U.S.–Canadian relations. To participate, send an abstract (150 words) to Professor Charles Doran, Program Chair, ACSUS '87, Canadian Center, School of Advanced International Studies, Johns Hopkins University, 1740 Mass. Ave., NW, Washington, D.C. 20036. A copy of the abstract should also be sent to ACSUS, One Dupont Circle, Suite 620, Washington, D.C. 20036. The deadline is February 1, 1987.

The National Center for Food and Agricultural Policy at Resources for the Future announces its 1987 Leadership Development Program. The four-week program offers up to 25 fellowships to qualified young professionals from business, government, education, farming, and agricultural organizations. Each Fellow carries out a self-directed study of a policy topic, as well as attending seminars and workshops, plus meetings with key decision makers. For full information and application materials, contact John J. Kornacki, National Center for Food and Agricultural Policy, Resources for the Future, 1616 P Street, NW, Washington, D.C. (telephone 202+328-5056).

American Statistical Association/NSF/Census Bureau Research Fellowships and Associateships beginning around September 1, 1987; positions at Census Bureau for one year or shorter period. General areas for research are economic measurement and analysis, social and demographic studies, and statistical methodology and computing. Requirements: for Fellows, Ph.D. and research record in relevant field (economics, statistics, demography, sociology); for Associates, at least two years graduate study in relevant field. Salaries are commensurate with qualifications and experience; fringe benefits and travel allowance are also provided. Apply by January 1, 1987 for Fellows; February 15, 1987 for Associates. For full information, contact Dr. Arnold Reznek, Coordinator, ASA/NSF/Census Research Program, RM 3524-3, SRD, Bureau of the Census, Washington, D.C. 20233 (telephone 301+736-3846).

The Law and Economics Center of the George Mason University School of Law (formerly at Emory) administers its Law Institute for Economists in May or June, 1987. The two-week course provides an intensive introduction to the American legal system, its institutions, research methods, and procedures. Participants receive full tuition, private rooms, most meals, and books. For exact dates and application forms, contact Assistant Dean Marc Hoberman, George Mason University, Law School/LEC, 3401 N. Fairfax Drive, Arlington, VA 22201. The deadline is in January, 1987.

Call for Papers: The 1987 meeting of the History of Economics Society will be held June 20-22 at the Harvard Business School Campus in Boston. Proposals to deliver papers and to be a discussant should be sent no later than December 30, 1986 to President-Elect Donald A. Walker, Department of Economics, Indiana University of Pennsylvania, Indiana, PA 15705.

Economists who are strongly oriented toward the humanities, who use humanistic methods in their research, and who will be participating in meetings held outside the United States, Mexico, and Canada that are concerned with the humanistic aspects of their discipline are eligible to apply for small travel grants of the American Council of Learned Societies. Financial assistance is limited to airfare between major commercial airports and will not exceed one-half of projected economy-class fare. Social scientists and legal scholars who specialize in the history or philosophy of their disciplines are eligible if the meeting they wish to attend is so oriented. Applicants must hold a Ph.D. degree or its equivalent, and must be citizens or permanent residents of the United States. To be eligible, proposed meetings must be broadly international in sponsorship or participation, or both. The deadlines for application to be received in the ACLS office are: for meetings scheduled between July and October, March 1; for meetings scheduled between November and February, July 1; for meetings scheduled between March and June, November 1. Please request application forms by writing directly to the ACLS (Attention: Travel Grant Program), 228 East 45 Street, New York, NY 10017, setting forth the name, dates, place, and sponsorship of the meeting, as well as a brief statement describing the nature of your proposed role in the meeting.

Deaths

Harold Davey, professor emeritus of economics, Iowa State University, April 19, 1986.

Martin P. Oettinger, senior lecturer, department of economics, University of California-Davis, April 15, 1986.

Daniel H. Saks, professor of educational policy, George Peabody College for Teachers, and professor of economics, Vanderbilt University, January 14, 1986.

Retirement

Edward G. Emerling: professor emeritus of economics, St. Bonaventure University, September 1986.

Daniel M. Holland: Sloan School of Management, Massachusetts Institute of Technology, June 30, 1986.

Roland T. Mullins: professor of economics and finance, Arkansas State University, May 1986.

Richard J. Solie: professor of economics, University of Alaska-Fairbanks, January 1986.

James R. Street: Rutgers-The State University, June 30, 1986.

Anthony M. Tang: professor of economics, Vanderbilt University, July 1, 1986.

Foreign Scholars

Ingo Boebel, University Erlangen-Nurnberg: visiting associate professor, Center for Political Economy, Washington University, January-April 1987.

Ben-Zion Zilberfarb, Bar-Ilan University: visiting professor of economics, Vanderbilt University, August 1, 1985.

Promotions

Christine E. Amsler: assistant professor of economics, Michigan State University, September 1986.

Kathryn H. Anderson: associate professor of economics, Vanderbilt University, September 1, 1986.

David L. Baumer: associate professor, North Carolina State University, August 16, 1986.

Vivian Bull: professor of economics, Drew University, September 1986.

Richard V. Burkhauser: professor of economics, Vanderbilt University, September 1, 1985.

Christine M. Cumming: research officer and senior economist, financial markets department, Federal Reserve Bank of New York, January 1, 1986.

Fred Curtis: associate professor of economics, Drew University, January 1986.

John C. Dutton: associate professor, North Carolina State University, August 16, 1986.

John R. Finlay: associate professor of economics, Wilfrid Laurier University, July 1, 1986.

Ronald C. Fisher: professor of economics, Michigan State University, July 1986.

Jose Garcia-Medrano: vice president and economist, Bank of America, June 1, 1986.

John H. Goddeeris: associate professor of economics, Michigan State University, July 1986.

Melvin L. Greenhut: professor of economics, Texas A&M University, June 1986.

Willene A. Johnson: manager, foreign exchange department, Federal Reserve Bank of New York, July 1, 1986.

Muhammed Kabir: associate professor of economics, University of New Brunswick in Saint John, July 1, 1986.

Charles R. Knoeber: professor, North Carolina State University, August 16, 1986.

Terrence J. Levesque: associate professor of economics, Wilfrid Laurier University, July 1, 1986.

Douglas J. McCready: professor of economics, Wilfrid Laurier University, July 1, 1986.

Esfandiar Maasoumi: professor of economics, Indiana University-Bloomington, July 1, 1986.

Allan S. Mandel: assistant director, Financial Assistance Program, U.S. Small Business Administration, December 1985.

Paul L. Menchik: professor of economics, Michigan State University, July 1986.

Joseph Pelzman: professor of economics, George Washington University, July 1, 1986.

J. Rex Pulley: professor of economics, Arkansas State University, August 1986.

James F. Rodgers: director, Center for Health Policy Research, American Medical Association, November 1985.

Rosalind S. Seneca: professor of economics, Drew University, September 1986.

Yeung-Nan Shieh: associate professor of economics, University of Alaska-Fairbanks, September 1986.

F. Bruce Simmons III: associate professor, department of management, University of Akron, September 1, 1986.

Steven M. Wiggins: associate professor, department of economics, and Center for Education and Research in Free Enterprise, Texas A&M University, January 1986.

David E. Wildasin: professor of economics, Indiana University-Bloomington, July 1, 1986.

Administrative Appointments

Jean W. Adams: acting associate dean, College of Sciences and Humanities, Iowa State University, 1986–87.

Carson W. Bays: chair, department of economics, East Carolina University, August 20, 1986.

Otis W. Gilley: head, department of economics, University of Alaska-Fairbanks, July 1986.

Robert W. Jolly: assistant dean, College of Agriculture, and assistant director, Agricultural Experiment Station, Iowa State University.

Henry G. Manne, Emory University: dean, School of Law, and director, Law and Economics Center, George Mason University, July 1, 1986.

S. Mestelman: chairman, department of economics, McMaster University, July 1, 1986.

Larry Peppers, Creighton University: dean, School of Commerce, Economics, and Politics, and professor of economics, Washington and Lee University, July 1, 1986.

Frank A. Sloan: chairman, department of economics, Vanderbilt University, June 1, 1986.

Robert Stuart: chairman, department of economics, Rutgers-The State University, July 1, 1984.

New Appointments

John Ambrose, Ohio State University: visiting lecturer, department of economics, University of Michigan, September 1, 1986.

Richard T. Baillie, University of Birmingham, England: associate professor of economics, Michigan State University, September 1986.

Michael R. Baye: assistant professor, department of economics, Texas A&M University, September 1985.

Clive Bell: professor of economics, Vanderbilt University, August 1, 1986.

Linda Bell: economist, domestic research department, Business Conditions Division, Federal Reserve Bank of New York, November 4, 1985.

Carole Biewener, University of Massachusetts-Amherst: visiting assistant professor of economics, Washington and Lee University, July 1, 1986.

David Cushman: visiting assistant professor of economics, Vanderbilt University, January 1986–June 1987.

Lawrence R. Dale, West Texas State University: associate professor of economics, Arkansas State University, August 1986.

Gary Arthur Dymski, The Brookings Institution: assistant professor, University of Southern California, fall 1986.

David W. Emmons: economist, department of medical practice, American Medical Association, May 1986.

Robert C. Feenstra: associate professor, department of economics, University of California-Davis, July 1, 1986.

James D. Gaisford: lecturer, department of economics, Wilfrid Laurier University, July 1, 1986.

Ira A. Gang, Cornell University: assistant professor of economics, Rutgers-The State University, September 1986.

Dermot J. Hayes: adjunct assistant professor, department of economics, Iowa State University, March 1, 1986.

Raj K. Chhikara: assistant professor, department of economics, Iowa State University, March 1, 1986.

Alvin E. Headen, Jr., American Medical Association: assistant professor, North Carolina State University, August 16, 1986.

Mark Herander: visiting associate professor of economics, Vanderbilt University, 1986–87.

James B. Hunt, Jr.: adjunct professor, North Carolina State University, July 1, 1986–June 30, 1987.

George G. Judge, University of Illinois: professor, department of agricultural and resource economics, University of California-Berkeley, July 1, 1986.

Avery Katz, Harvard University: assistant professor, department of economics, and adjunct assistant professor, Law School, University of Michigan, September 1, 1986.

Ronald D. Kneebone: lecturer, department of economics, Wilfrid Laurier University, July 1, 1986.

James Leitzel: assistant professor of economics, Vanderbilt University, August 1, 1986.

Stanley J. Liebowitz, University of Chicago: associate professor, North Carolina State University, August 16, 1986.

Robert R. Logan, University of Iowa: assistant professor of economics, University of Alaska-Fairbanks, September 1986.

Gabriel A. Lozada: assistant professor, Texas A&M University, September 1986.

David A. Malueg, University of Iowa: economist, Economic Regulatory Section, Antitrust Division, U.S. Department of Justice, June 1986.

Akbar Marvasti, Louisiana State University: visiting assistant professor of economics, East Carolina University, August 20, 1986.

Edwina A. Masson, Michigan State University: assistant professor of economics, Rutgers-The State University, September 1986.

Janet Mitchell, Northwestern University: assistant professor, University of Southern California, fall 1986.

Edward B. Montgomery, Carnegie-Mellon University: associate professor of economics, Michigan State University, September 1986.

Swati Mukerjee, Boston University: assistant professor, department of economics, Trinity University, August 1986.

Samuel L. Myers, Jr., Graduate School of Public and International Affairs, University of Pittsburgh: professor of Afro-American studies and economics, University of Maryland, July 1, 1986.

Randall Parker, University of Kentucky: assistant professor of economics, East Carolina University, August 20, 1986.

Peter Rappoport: economist, domestic research department, Business Conditions Division, Federal Reserve Bank of New York, December 30, 1985.

James Reitzes: economist, international research department, Industrial Economies Division, Federal Reserve Bank of New York, December 4, 1985.

Richard Roehl, University of Michigan-Dearborn: visiting professor, department of economics, University of Michigan, September 1, 1986.

Clifford Russell: professor of economics, Vanderbilt University, August 1, 1985.

Todd Sandler, University of South Carolina: professor, department of economics, Iowa State University, June 1986.

Mohammed A. Shibli, Boston University: assistant professor, department of business administration, Stonehill College, September 1986.

Bernard Smith, Yale University: instructor of economics, Drew University, September 1986.

Michael Smitka, Yale University: assistant professor of economics, Washington and Lee University, July 1, 1986.

Janet G. Stotsky, Stanford University: assistant professor of economics, Rutgers-The State University, September 1986.

Russell Stout, Jr.: visiting associate professor, department of economics, Iowa State University, December 15, 1985.

Pravin K. Trivedi, Australian National University: professor of economics, Indiana University-Bloomington, January 1987.

Kenneth Weiller: economist, international research department, Industrial Economies Division, Federal Reserve Bank of New York, February 19, 1986.

Eugene N. White, University of Illinois-Urbana: associate professor of economics, Rutgers-The State University, July 1, 1986.

Richard J. Wilke: economist, Health Resource Analysis, American Medical Association, September 1985.

Glenn A. Woroch, University of Rochester: senior member technical staff, GTE Laboratories, Inc., May 1986.

Lester Zeager, University of Pittsburgh: assistant professor of economics, East Carolina University, August 20, 1986.

Resignations

Lourdes Beneria, Rutgers-The State University: Cornell University, June 30, 1985.

Stephen Bronars, Texas A&M University, September 1986.

Peter Calkins, Iowa State University: Laval University, February 28, 1986.

Roger Dahlgran, Iowa State University: University of Arizona-Tucson, October 1985.

Steven L. Green, Vanderbilt University: Baylor University, August 1986.

Woo Sik Jung, Vanderbilt University: University of Kansas, August 1985.

Marc A. Miles, Rutgers-The State University, June 30, 1986.

Harris Schlesinger, Vanderbilt University: University of Illinois-Urbana.

NOTE TO DEPARTMENTAL SECRETARIES AND EXECUTIVE OFFICERS

When sending information to the *Review* for inclusion in the Notes Section, use the following style:

A. Please use the following categories (please—do not send public relation releases):

1—Deaths
2—Retirements
3—Foreign Scholars (visiting the USA or Canada)
4—Promotions
5—Administrative Appointments
6—New Appointments
7—Leaves for Special Appointments (NOT Sabbaticals)
8—Resignations
9—Miscellaneous

B. Please give the name of the individual (SMITH, Jane W.), her present place of employment or enrollment: her new title (if any), new institution and the date at which the change will occur.

C. Type each item on a separate 3×5 card.

D. The closing dates for each issue are as follows: *March*, October 15; *June*, January 15; *September*, April 15; *December*, July 15.

All items and information should be sent to the Editorial Office, *American Economic Review*, 209 Nassau Street, Princeton, NJ 08542-4607.

EIGHTY-THIRD LIST OF DOCTORAL DISSERTATIONS IN ECONOMICS IN AMERICAN UNIVERSITIES AND COLLEGES

The present list specifies doctoral degrees conferred during the year terminating June 1986.

General Economics; including Economic Theory, History of Thought, Methodology, Economic History, and Economic Systems

SIMON P. ANDERSON, Ph.D. Queen's 1985. Three essays on address models of value theory.

MARCELLUS W. ANDREWS, Ph.D. Yale 1986. Competition and macroeconomics: Some implications of Schumpeter's theory of development for macroeconomics.

NASIR H. ASSAR, Ph.D. West Virginia 1985. Impact of declining saving rate on economic aggregates: A comparative study.

AVNER BAR-ILAN, Ph.D. Princeton 1985. Optimal decision rules and fixed transaction costs.

HUMBERTO BARRETO, Ph.D. North Carolina-Chapel Hill 1985. The entrepreneur in microeconomic theory —disappearance and explanation.

SUSHIL BIKHCHANDANI, Ph.D. Stanford 1986. Market games with few traders.

ORN B. BODVARSSON, Ph.D. Simon Fraser 1986. The economics of employee monitoring: A sequential measurement approach.

J. RODY BORG, Ph.D. North Carolina-Chapel Hill 1986. The profitability of acquisitions: 1905–30.

JOHN H. BOYD III, Ph.D. Indiana-Bloomington 1986. Preferences, technology, and dynamic equilibria.

DOUGLAS M. BROWN, Ph.D. Colorado State 1985. The Budapest School: A post-Marxian defense of the mixed economy.

JAMES D. CAVALLO, Ph.D. New York 1986. Economic inequality and the intergenerational transfer of income-earning potential.

SUCHAN CHAE, Ph.D. Pennsylvania 1985. Three essays in economic theory.

SUBIR CHAKRABARTI, Ph.D. Iowa 1985. Refinements of some core-like solution concepts and an application to oligopoly theory.

IN-KOO CHO, Ph.D. Princeton 1986. Refinement of sequential equilibrium: Theory and application.

YOUNG BACK CHOI, Ph.D. Michigan 1986. Decision making and economic behavior under uncertainty: A paradigmatic approach.

GEORGE D. CHOKSY, Ph.D. Tennessee-Knoxville 1986. An epistemological analysis of the economic writings of Sir Dudley North.

HARVEY S. CUTLER, Ph.D. Washington 1985. Aggregate supply and demand disturbances and the natural rate of output.

CHARLES A. DE BARTOLOME, Ph.D. Pennsylvania 1985. The public provision of private goods.

FANNY DEMERS, Ph.D. Johns Hopkins 1986. Bayesian learning as a propagating mechanism in a dynamic general equilibrium model of business cycles.

MICHAEL DEMERS, Ph.D. Johns Hopkins 1986. Essays in the economics of uncertainty.

MICHAEL B. DEVEREUX, Ph.D. Queen's 1986. Inflation, exchange rates, and economic policy.

JANNETTE O. DOMINGO, Ph.D. Columbia 1986. The role of imported labor in Virgin Islands economic history.

B. KELLY EAKIN, Ph.D. North Carolina-Chapel Hill 1985. Estimating allocative inefficiency with a non-minimum cost function.

SHARON J. ERENBURG, Ph.D. Illinois-Urbana 1986. The effects of permanent and transitory variations on aggregate supply and price.

IAN L. GALE, Ph.D. Princeton 1986. Essays on asymmetric information in asset trading models.

ROBERT S. GIBBONS, Ph.D. Stanford 1986. Essays on labor markets and internal organization.

JOYCE G. GLEASON, Ph.D. Nebraska-Lincoln 1986. Economic models of time in learning.

FARUK GÜL, Ph.D. Princeton 1986. Two essays on the bargaining foundations of value theory.

KENT S. HALL, Ph.D. George Washington 1986. Business failures: Economic determinants and potential effects on the macroeconomy.

WARD A. HANSON, Ph.D. Stanford 1985. Dynamic pricing of competing technological systems subject to decreasing costs.

WILLIAM HENRY, Ph.D. Pittsburgh 1985. The evolution of monetary thought in the United States, 1865–1929.

JOHN J. HISNANICK, Ph.D. West Virginia 1986. The local dynamic stability of a general equilibrium system with varying speed of adjustment coefficients.

DAVID J. HOAAS, Ph.D. Duke 1986. The political economy of tax-based incomes policy.

ROLAND HOKSBERGEN, Ph.D. Notre Dame 1986. On the character and importance of value judgements in economics, with an application to development project evaluation.

THOMAS J. HOLMES, Ph.D. Northwestern 1985. Price discrimination in imperfectly competitive markets.

JOHN F. HUGHES, Ph.D. Northwestern 1986. Industrial market structure and aggregate output fluctuations:

The Means thesis reconsidered.

CORWIN C. HULBERT, Ph.D. Nebraska-Lincoln 1986. Choice and market processes: An Austrian analysis.

HOSSEIN KAMAREI, Ph.D. Indiana-Bloomington 1985. The construction and experimental testing of general equilibrium models with monopolistic interaction.

KYUN KIM, Ph.D. Duke 1986. Equilibrium business cycle theory in historical perspective.

GERASIMOS D. KOUMATOS, Ph.D. Pennsylvania 1985. The principal-agent problem with multidimensional agents.

PRAVEEN KUMAR, Ph.D. Stanford 1985. Essays on intertemporal incentives and signaling.

PAUL H. KUPIEC, Ph.D. Pennsylvania 1985. Microfoundations of systematic risk.

JERRY LANGIN-HOOPER, Ph.D. Rutgers 1985. A static, partial equilibrium, rate-of-return based-rates model of the regulation of natural monopolies: A revision of the conventional Averch and Johnson analysis.

CARLOS I. S. LEAL, Ph.D. Princeton 1986. Topics in noncooperative game theory: Hyperstability of Nash equilibria, Brown-von Neumann fictitious play and principal's problem.

JON T. LEVEVER, Ph.D. Northwestern 1986. Capital utilization: A theoretical and empirical analysis of the duration of production.

BARTON L. LIPMAN, Ph.D. Michigan 1985. Delaying or deterring entry: A game-theoretic analysis.

DOUGLAS A. MCMANUS, Ph.D. Pennsylvania 1985. Essays in the theory of rational expectations in microeconomics and econometrics.

FRANK A MACHOVEC, Ph.D. New York 1986. The destruction of competition theory: The perfectly competitive model and beyond.

ABELHAMID MAHBOB, Ph.D. Southern California 1985. North-South trade and the growth of the South.

GEORGE J. MAILATH, Ph.D. Princeton 1985. Separating equilibria in signaling games: Incentive compatibility, existence with simultaneous signaling, welfare, and convergence.

RODOLFO E. MANUELLI, Ph.D. Minnesota 1986. Topics in intertemporal economics.

LARRY J. MOORE, Ph.D. Indiana-Bloomington 1985. The Wicksell-Keynes price model.

PETER S. NICULESCU, Ph.D. Yale 1985. Indivisibilities, commitments, and expected utility theory.

JOHN V. NYE, Ph.D. Northwestern 1985. Firm size and economic backwardness: A new look at the French industrialization debate.

JAY Y. OHM, Ph.D. Northwestern 1986. Differential game approach to the complementary monopoly problem.

BENOIT-M. PAPILLON, Ph.D. Northwestern 1986. Montreal's growth and economic changes in Quebec Province, 1851–1911.

JOHN E. PARSONS, Ph.D. Northwestern 1986. Two essays on the economics of sequential markets with imperfect information.

STUART W. PASSMORE, Ph.D. Michigan 1985. Essays in microeconomics: Can unions be barriers to entry? Estimating elasticity of demand for goods priced by time of day; Ramsey optimal price and capacity with uncertainty.

JAMES D. PECK, Ph.D. Pennsylvania 1985. Essays in intertemporal economic theory.

MIKKO T. PUHAKKA, Ph.D. Pennsylvania 1985. On the foundations of disequilibrium macroeconomics.

SAAD ALI RABHAN, Ph.D. Georgetown 1986. On the economics of exhaustible resources.

WILLIAM W. ROSS, Ph.D. Georgetown 1986. The effects of financial disturbances and government policies in a growing open economy.

SALVATOR L. M. G. RUGIMBANA, Ph.D. Catholic 1986. Generation of expectations variables and their use in testing the properties of demand functions using the almost ideal system.

TATSUYOSHI SAIJO, Ph.D. Minnesota 1985. Nash implementation theory.

BARBARA N. SANDS, Ph.D. Washington 1985. An investigation into the nature and extent of the market in Shanxi Province, China, 1928 to 1945.

DAVID SCHARFSTEIN, Ph.D. MIT 1986. Market competition and incentives.

AUSTIN L. SCHMITT, Ph.D. George Washington 1986. The roots of disequilibrium theory.

PAUL J. SCHOOFS, Ph.D. North Carolina-Chapel Hill 1985. A critique of some recent contributions to demand theory.

JOHN SCHULTZ, Ph.D. South Carolina 1985. Essays on the history and philosophy of economic science.

JANET A. SEIZ, Ph.D. Duke 1986. Business cycle theory and the "Keynesian Revolution": A reexamination.

MOHAMMED A. SHIBLI, Ph.D. Boston 1986. Investment opportunities, household savings, and rate of return on investment.

MARTHA K. SHIELLS, Ph.D. Michigan 1985. Hours of work and shift work in the early industrial labor markets of Great Britain, the United States, and Japan.

BYUNG HEE SOH, Ph.D. Northwestern 1985. Economic fluctuations and government policy under democracy.

JEREMY STEIN, Ph.D. MIT 1986. Economic models with heterogeneously informed participants.

PETER A. STREUFERT, Ph.D. Stanford 1985. Dynamic allocation with consistent intergenerational benevolence.

RICHARD SZOSTAK, Ph.D. Northwestern 1985. The role of transportation in the Industrial Revolution: A comparison of eighteenth-century England and France.

Harutaka Takahashi, Ph.D. Rochester 1985. Characterization of optimal programs in infinite horizon economies.

Andrea Terzi, Ph.D. Rutgers 1986. Savings, expenditures, and the financial conditions of production: Two views.

Robert G. Trevor, Ph.D. Princeton 1985. Rational expectations and economic modeling: Four essays in macroeconomics and international finance.

Stephen J. Turnbull, Ph.D. Stanford 1985. Durable mechanisms for bargaining with private information: Definition, existence, and characterization.

Carolyn Tuttle, Ph.D. Northwestern 1986. Children at work in the British Industrial Revolution.

Nancy L. Virts, Ph.D. California-Los Angeles 1985. Plantations, land tenure and efficiency in the postbellum South: The effects of emancipation on southern agriculture.

Marian W. D. Watts, Ph.D. Alabama 1986. Corporatism and Roman Catholic socioeconomic thought: An analysis of current issues.

Sergio R. d.C. Werlang, Ph.D. Princeton 1986. Common knowledge and the game theory.

R. White, Ph.D. American 1985. The disequilibrium effects of technological change on prices and the rate of profit.

Jennifer Wissink, Ph.D. Pennsylvania 1985. Essays in applied microeconomic theory.

Peter J. Wylie, Ph.D. Queen's 1986. Electrification and technological adaptation in Canadian manufacturing, 1900–29.

Economic Growth and Development; including Economic Planning Theory and Policy; Economic Fluctuation and Forecasting

Wellappuli A. T. Abeysekera, Ph.D. Cornell 1986. Improvement of irrigation system performance in Sri Lanka: The experience at Mahavilachiya Settlement.

Munir Z. Abu-Ghazaleh, Ph.D. Vanderbilt 1985. Economics of tourism in Jordan.

Chul Won Ahn, Ph.D. Vanderbilt 1985. Essays on comparative monetary policy regime analysis.

Jafar Alavi, Ph.D. Tennessee-Knoxville 1986. The export performance of less developed countries: The case of Iran's non-oil exports.

Abdul-Karim A. Amer, Ph.D. Colorado 1985. Economic planning in Yemen Arab Republic: The dependency problem.

Ignazio Angeloni, Ph.D. Pennsylvania 1985. Credit controls, bank behavior, and monetary policy: Essays on the Italian experience job.

Muhammad Anwar, Ph.D. New Hampshire 1985. A macroeconomic model for interest-free economies: An integrative study of Western and Islamic economic systems.

F. Arafat, Ph.D. American 1985. Development with unlimited supplies of imported labor: A case study of the kingdom of Saudi Arabia.

Ian R. Bain, Ph.D. Minnesota 1985. A theory of the secular and cyclical behavior of inventory stocks.

John H. Baskin III, Ph.D. Georgia 1986. A quarterly econometric model of the state of Georgia.

Raymond G. Batina, Ph.D. Minnesota 1985. An overlapping generations model with endogenous fertility and intergenerational transfers.

Carlos M. Benevides, Ph.D. Vanderbilt 1985. Income distribution in Brazil: 1970–80 compared.

Mark J. Bils, Ph.D. MIT 1985. Essays on the cyclical behavior of cost and price.

Henning Bohn, Ph.D. Stanford 1986. Financial markets and the macroeconomy.

Lenore H. Burckel, Ph.D. Wisconsin-Milwaukee 1986. Comparing and evaluating alternative measures of inflation expectations.

Wen-Hui Peter Cheng, Ph.D. Minnesota 1985. An educational planning model of growth and income distribution: A case study of Taiwan.

Hyung In Chin, Ph.D. SUNY-Binghamton 1986. The firm's dynamic adjustment behavior: An empirical study in the microeconomic level.

Munir Choudhary, Ph.D. Florida State 1986. Comparative effectiveness of stabilization policies with rational expectations: Developed versus developing countries.

Sonia Dahab, Ph.D. Yale 1985. The agricultural machinery and implement industry in Brazil: Its historical development and incentive activity.

Steve J. Davis, Ph.D. Brown 1986. Allocation disturbances, aggregate disturbances, and unemployment rate fluctuations.

Paul A. Dorosch, Ph.D. Stanford 1986. Linkages between the macroeconomy of Indonesia's food sector to 1949 and 1984.

J. Ehdaie, Ph.D. American 1985. A simple macroeconomics model of inflation: The case of Iran.

Atilio A. Elizagaray, Ph.D. Illinois-Urbana 1985. State intervention in the Argentine economy: The 1946–55 Pernoiut government.

John L. Fiedler, Ph.D. Vanderbilt 1986. Health and development in El Peten, Guatemala.

Raimundo C. Florin, Ph.D. Illinois-Urbana 1986. Financial history of Argentina: 1890–1910.

Jaime B. Gatica, Ph.D. Notre Dame 1986. De-industrialization and employment: The case of Chile, 1974–82.

Paul W. Glewwe, Ph.D. Stanford 1985. An analysis of income distribution and labor markets in Sri Lanka.

Benjamin Gomes-Casseres, Ph.D. Harvard 1986. Multinational ownership strategies.

E. Gordji-Bandpay, Ph.D. American 1985. The redistributive effects of government fiscal policy in Iran in 1970–80.

Joaquim J. M. Guilhoto, Ph.D. Illinois-Urbana 1986.

A model for economic planning and analysis for the Brazilian economy.

CYRILLUS HARINOWO, Ph.D. Vanderbilt 1985. Exchange rates and asset substitution in the Indonesian monetary sector.

ETHAN S. HARRIS, Ph.D. Columbia 1985. A reexamination of inventory buffer effect with disaggregated data.

JANE A. HORVATH, Ph.D. Connecticut 1986. The role of small scale manufacturing in Mexican development.

WEN-JEN HSIEH, Ph.D. Columbia 1985. Technology transfer: The case of Taiwan.

MOHAMED A. IBNOUF, Ph.D. Michigan State 1985. An economic analysis of the mechanized food production schemes in the central plains of the Sudan.

NARENDRA JADHAV, Ph.D. Indiana-Bloomington 1986. Macroeconomic investment management in LDCs: A social cost benefit approach.

S. M. YOUNUS JAFRI, Ph.D. Wayne State 1985. General equilibrium macroeconomic modeling: An application to Pakistan's sixth plan (1983–88).

JOSE C. JARILLO MOSSI, Ph.D. Harvard 1986. Entrepreneurship and growth: The strategic use of external resources.

STEVEN SHEI-PING KAN, Ph.D. Purdue 1985. Three essays on the Lucas supply function, partial adjustment of inventories, and testing of the permanent income hypothesis.

PETROS KAREKLAS, Ph.D. South Carolina 1986. A monetary analysis of a small developing economy: The case of Jamaica.

FRANK KRAMER, Ph.D. California-Berkeley 1986. Market incorporation and out-migration of the peasants of western Honduras.

B. KRISHNAMURTHI, Ph.D. American 1985. Public expenditure policy implications for output, prices, and balance of payments, and empirical study of the Indian economy.

MARK L. LANGWORTHY, Ph.D. Stanford 1986. The impacts of EC membership on the Portuguese dairy sector.

YOON JOO LEE, Ph.D. Catholic 1986. Effective rate of protection and factor substitution: A case study of Korea and the Ivory Coast.

CARLOS M. D.M. LOPES, Ph.D. Illinois-Urbana 1986. An evaluation of rural credit policies in Brazil.

JAMES P. MCCOY, Ph.D. North Carolina-Chapel Hill 1986. Bauxite processing in the Third World: An estimation of producer country benefits.

CATHERINE L. MCDEVITT, Ph.D. Rochester 1986. Real returns and inflation: Theory and evidence.

DUNCAN C. MCDOUGALL, Ph.D. Harvard 1986. Manufacturing strategy in diversified firms: Linkages between competitive strategy and manufacturing planning in the divisions.

BRUCE MCNEVIN, Ph.D. City-New York 1986. Business cycles and asymmetry.

EDWARD D. MARTIN, Ph.D. Cornell 1986. Resource mobilization, water allocation, and farmer organization in hill irrigation systems in Nepal.

LAURIE J. MAY, Ph.D. Pennsylvania 1985. The returns to schooling and the demand for education in Brazil.

PETER H. MAY, Ph.D. Cornell 1986. A modern tragedy of the non-commons: Agro-industrial change and equity in Brazil's Babassu Palm Zone.

ROBERT S. MILBRATH, Ph.D. Michigan 1986. Institutional development and capital accumulation in a complex of Basque worker cooperatives.

ARMANDO L. MONTENEGRO, Ph.D. New York 1985. Optimal economic growth for an economy with trade complementarity and indispensability of imports.

BARRY J. NAUGHTON, Ph.D. Yale 1986. Saving and investment in China: A macroeconomic analysis.

CHARLES N'CHO, Ph.D. Stanford 1985. A mathematical programming evaluation of alternative development strategies.

ALFRED N. NTOKO, Ph.D. SUNY-Binghamton 1986. Government policy and agricultural growth: The case of Cameroon.

PHILIP A. NUETZEL, Ph.D. Northwestern 1986. Inflation and capital accumulation in models of optimizing behavior.

KEVIN J. O'CONNOR, Ph.D. 1986. An economic and legal analysis of predation.

INSHIK OH, Ph.D. Vanderbilt 1985. The life cycle hypothesis of aggregate savings: Static and dynamic analysis of simulation model.

JOSEPH O. ONWE, Ph.D. Wisconsin-Milwaukee 1986. Labor absorption in the services sector: An analysis with reference to transport services in Nigeria.

NEHEMIAH E. OSORO, Ph.D. Illinois-Urbana 1985. The elasticity and bouyancy of the tax system of Tanzania, 1968–81: An empirical analysis.

SRI-BINTANG PAMUNGKAS, Ph.D. Iowa State 1984. A medium-term multisectoral dynamic simulation model of the Indonesian economy.

J. PETERSON, Ph.D. American 1985. The formation of expectations and the demand for capital.

THOMAS C. PINCKNEY III, Ph.D. Stanford 1986. Production instability and food security in Kenya: Measuring tradeoffs between government objectives.

SUSAN G. POWERS, Ph.D. SUNY-Binghamton 1986. Cyclical variation in capital stock: Implications for multifactor productivity.

MUNIR QUDDUS, Ph.D. Vanderbilt 1985. Rational speculative bubbles in the food market: A study of the 1974 famine in Bangladesh.

JAMES E. RAUCH, Ph.D. Yale 1985. Dynamic interaction of foreign trade, technology, and capital with economic development.

LAKSHMI K. RAUT, Ph.D. Yale 1985. Three essays on intertemporal economic development.

PEDRO N. SAENZ, Ph.D. Georgetown 1986. Endowments, migration, and economic development: An adaptive modeling approach.

KENICHI SAKAKIBARA, Ph.D. Minnesota 1986. Expectations and the nonneutrality of money: A legal restriction theory of business cycles.

MARIO SALAZAR, Ph.D. Iowa State 1984. An multicommodity equilibrium approach to welfare analysis of market interventions in the Costa Rican agricultural sector.

T. MISHA SARKOVIC, Ph.D. Florida State 1986. Direct foreign investment and joint-venture agreements in Yugoslavia: A microeconomic model.

JAMES L. SEALE, JR., Ph.D. Michigan State 1985. Fixed effect Cobb-Douglas production functions for floor tile firms, Fayoum and Kalyubiya, Egypt, 1981–83.

EXLEY SILUMBU, Ph.D. Notre Dame 1985. Commodity marketing and economic restructuring: The role of smallholder agriculture in Malawi.

RUSSELL E. SMITH, Ph.D. Illinois-Urbana 1985. Wage policy and nominal wages in Brazil.

DAVID F. SPIGELMAN, Ph.D. Stanford 1986. Macroeconomic instability of the less developed country economy when bank credit is rationed.

KENNETH G. STEWART, Ph.D. Michigan 1986. Estimating a rational expectations business cycle model: Methodology and application.

MOHAMMED TAHIR, Ph.D. Colorado 1985. The effectiveness of stabilization policies in an open economy: An application of the vector autoregressive model.

TESFAYE TEKLU, Ph.D. Iowa State 1984. An economic analysis of joint resource and income use practices among the Ethiopian farm households: An application of production household model.

LUIS TELLEZ-KUENZLER, Ph.D. MIT 1986. Essays on real and financial aspects of an open economy: The Mexican case.

JOSE R. O. VERGOLINO, Ph.D. Illinois-Urbana 1986. Unequal regional development: The case of the Brazilian Amazon.

DARAWAN VIRUNHAPHOL, Ph.D. SUNY-Binghamton 1986. A macroeconometric model of Thailand.

IVAN L. WEINEL, Ph.D. Nebraska-Lincoln 1985. Financial instability and business fluctuations.

CHANDRASEKERA M. WIJAYARATNA, Ph.D. Cornell 1986. Assessing irrigation system performance: A methodological study with application to Gal Oya, Scheme, Sri Lanka.

KOON FOO WONG, Ph.D. Southern California 1985. Three related aspects of married women's labor supply decisions: An application to Malaysia.

ABDUL A. B. YAACOB, Ph.D. Maryland-College Park 1986. Islamic banking and economic development.

HUIZHONG ZHOU, Ph.D. Northwestern 1986. Dynamic effects of incentive schemes in economic planning.

Economic Statistics; including Econometric Methods, Economic and Social Accounting

SAMUEL K. ANDOH, Ph.D. New York 1986. The demand for cocoa: Explaining low-income elasticity of demand in the high-income countries.

FRANK J. ATKINS, Ph.D. Queen's 1986. A frequency domain analysis of the long-run relation between nominal interest rates and inflation in Canada and the United States.

KYLE W. BAGWELL, Ph.D. Stanford 1986. Signaling, entry, and repeat business.

PAUL W. BAUER, Ph.D. North Carolina-Chapel Hill 1985. An analysis of multiproduct technology and efficiency using the joint cost function and panel data: An application to the U.S. airline industry.

DALE S. BREMMER, Ph.D. Texas A&M 1985. An econometric model of petroleum drilling in Texas.

KUI-YIN CHEUNG, Ph.D. Washington 1986. Signal extraction and two illustrations of the quantity theory of money: Application of Kalman filtering and smoothing.

HADI DADASHI-NOBARI, Ph.D. Washington 1985. Productivity growth slowdown, and *R&D* expenditures: Inflation, variation in input prices, in U.S. manufacturing.

MICHAEL A. DIMMITT, Ph.D. SUNY-Albany 1986. Time and space modeling: Asymptotically efficient instrumental variables estimators of export base modifiers.

HANS G. EHRBAR, Ph.D. Michigan 1985. The duality between unbiased estimation/prediction in the linear model and the associated least square problem, and the theory of mean-variance efficient portfolio: Two applications of quadratic minimization, G-inverses, and deficiency matrices.

ROBERT GILLETTE, Ph.D. Maryland 1985. A partially adaptive panel data estimator: A Monte Carlo simulation and an application to the forward foreign exchange market.

BASILE GOUNGETAS, Ph.D. Missouri-Columbia 1986. The impact of household size and composition on food consumption: An analysis of the food stamp program parameters using the NFCS, 1977–78.

STEVEN M. HOTOPP, Ph.D. Illinois-Urbana 1985. Practical methods in multivariate time-series analysis.

JAMIE F. HOWELL, Ph.D. Columbia 1985. The implications of a sample selectivity framework on the estimation of hedonic asking and offer rent equations.

FORREST E. HUFFMAN, Ph.D. South Carolina 1985. An integrated comprehensive model of economic base analysis with an application to the Columbia, South Carolina metropolitan statistical area.

RICHARD H. JEFFERIS, JR., Ph.D. Stanford 1985. Two essays on uncertainty.

JIN-HO JEONG, Ph.D. Indiana-Bloomington 1985. Lower risk reduced-form estimation and forecasting based on finite observations and uncertain structural specifications in a linear simultaneous equation model.

IL S. JUN, Ph.D. Tennessee-Knoxville 1986. Intertemporal and interspatial productivity comparisons of manufacturing industries for 53 SMSAs, 1963–78.

BYUNG-SUH KANG, Ph.D. South Carolina 1985. A comparative study of scheduling algorithms for flowshops with no intermediate storage.

MARIANO KLINGE-LOY, Ph.D. Rutgers 1986. Estimating factor-augmenting technical change: A Bayesian gradual switching regression approach.

RICHARD A. KOSS, Ph.D. Pennsylvania 1985. Some implications of the Kalman filter for econometric theory and practice.

SUBAL C. KUMBHAKAR, Ph.D. Southern California 1986. Estimation of inefficiency in production: Methodological and empirical issues.

YOUNG Y. LEE, Ph.D. Missouri-Columbia 1985. Implicit prices of and household demand for nutrient availability and variety indices: Hedonic food-group-cost-function approach.

JIA-YUAN LIN, Ph.D. Wayne State 1986. Estimation of disequilibrium models when the dependent variable is a Tobit variable.

JOSEPH LYNCH, Ph.D. South Carolina 1985. South Carolina economic indicators: A time-series approach.

SYED F. MAHMUD, Ph.D. McMaster 1986. Supply-side effects of monetary policy.

ROSA L. MATZKIN, Ph.D. Minnesota 1986. Mathematical and statistical inferences from demand data.

SHIGETAKA MIYAZAKI, Ph.D. Illinois-Urbana 1985. Risk comparison of LS, ML, and Stein estimators under various assumptions.

NADER NAZMI, Ph.D. Illinois-Urbana 1985. Time-series forecasting and leading indicators of economic activity.

CHRISTOPHER J. NICOL, Ph.D. Queen's 1986. Estimating higher-order flexible functional form demand systems with Canadian cross-section micro data.

MICHAEL B. O'BRIEN, Ph.D. South Carolina 1986. An analysis of measures of geographical cost of living differences in the United States.

PETER OCHSHORN, Ph.D. Michigan 1986. Regression specification in error tests.

PIERRE PERRON, Ph.D. Yale 1986. Hypothesis testing in time-series regression with a unit root.

SIMON A. T. B. POWER, Ph.D. Western Ontario 1986. Single-equation errors-in-variables estimation of rational expectations models.

SUBARNA K. SAMANTA, Ph.D. Southern Methodist 1985. Applications of Stein-Rule estimation to combination forecasting.

ANTHONY J. SMALL, Ph.D. New York 1986. Technique diffusion and input-output coefficient updating.

DEAN G. SMITH, Ph.D. Texas A&M 1985. A regional forecasting model of the hospital sector.

CHEIN-HSING SUNG, Ph.D. Wayne State 1986. Estimation of a modified linear spline regression model: Theory and application.

KEUK JE SUNG, Ph.D. Northwestern 1986. Product differentiation and entry deterrence.

CHRISTIE L. TEIGLAND, Ph.D. SUNY-Albany 1986. An econometric analysis of ASA-NBER survey data on expectations: 1968–81.

INSAN F. TUNALI, Ph.D. Wisconsin-Madison 1985. Migration, earnings, and selectivity: From theory to fact, evidence from Turkey, 1963–73.

MARY E. WALKER, Ph.D. Rice 1986. Issues concerning systems of demand equations.

GANG YI, Ph.D. Illinois-Urbana 1986. Stein estimators in model selection.

JAMAL E. ZAYID, Ph.D. Indiana-Bloomington 1986. The foreign sector and economic development: An analysis and simulations of an econometric model of the Sudan.

Monetary and Fiscal Theory, Policy and Institutions

SHAGHIL AHMED, Ph.D. Rochester 1985. Essays on intertemporal substitution and wage contract models of business fluctuations.

YONG-HO BAEK, Ph.D. SUNY-Albany 1986. Maximum-likelihood estimation of the consumption-based capital-asset pricing model.

ROBERT BARSKY, Ph.D. MIT 1986. Three interest rate paradoxes.

ABDULLA M. BASODAN, Ph.D. Washington 1985. Estimating private demand for a publicly supplied good: An hedonic price model application.

AHMET BAYTAS, Ph.D. City-New York 1986. Interest elasticity of money demand in the new financial environment.

ROLAND BENABOU, Ph.D. MIT 1986. Optimal price dynamics speculation and search under inflation.

WILLIAM C. BOSWORTH, Ph.D. Connecticut 1985. Unused credit as a substitute for money.

ROBERT A. BROWN, Ph.D. Northwestern 1986. The existence and effects of the information-based Keynesian speculative equilibrium hypothesis of asset pricing.

ROBERT D. BRUMBAUGH, JR., Ph.D. George Washington 1986. The determinants of failure and insolvency for federally insured thrift institutions and the determinants of federal savings and loan insurance corporation costs.

ANTHONY R. CHAN, Maryland 1986. A look at the simple-sum monetary aggregates versus other alternatives.

MOHAMMED M. CHAUDHURY, Ph.D. Simon Fraser 1986. On testing the Black-Scholes option pricing model.

JOSEPH CHENG, Ph.D. SUNY-Binghamton 1986. An econometric analysis of the effects of the flat income tax on household savings.

Kyungbae Chung, Ph.D. Pittsburgh 1986. Optimal monetary rules with an endogenous money supply.

John P. Cochran, Ph.D. Colorado 1985. Monetary theory, price theory, and capital theory: The Keynes-Hayek debate and the current policy effectiveness debate.

Roger S. Cohen, Ph.D. North Carolina-Chapel Hill 1985. A simulation analysis of voting to determine the allocative impacts of general fund financing and its alternatives.

J. Pedro D.A. Arroja, Ph.D. Carleton 1986. Monetary gradualism in Canada (1975–82): An evaluation.

S. Devadoss, Ph.D. Iowa State 1985. The impacts of monetary policies on U.S. agriculture.

John V. Eagan, Ph.D. Georgia State 1986. Essays on taxation, investment, and exhaustible resources.

John H. Y. Edwards, Ph.D. Maryland 1985. Congestion function specification and the publicness of goods provided by local governments.

Philip R. Erlanger, Ph.D. Pennsylvania 1985. A contingent claims model of monetary policy.

Miguel Faig-Aumalle, Ph.D. Stanford 1986. Optimal taxation of money balances.

Frederick G. Floss, Ph.D. SUNY-Buffalo 1986. The general equilibrium equity and efficiency effects of the economic recovery act of 1981 on heterogeneous labor: A simulation approach.

Gladys B. Foster, Ph.D. Colorado-Boulder 1986. The endogeneity of money and Keynes' *General Theory*.

Jean A. Gauger, Ph.D. Iowa State 1984. Three essays on the neutrality of anticipated money growth.

William A. Gibson, Ph.D. Simon Fraser 1986. Money, debt, and taxation in a model of government finance.

Jose M. Gonzalez-Paramo, Ph.D. Columbia 1986. On partial factor mobility and the short-run incidence of selective capital income taxes.

Marc Halpern, Ph.D. City-New York 1985. Call options on stock indexes.

Richard X. Hanson, Ph.D. California-Los Angeles 1986. Antecedents of the 1929 recession.

Douglas Holtz-Eakin, Ph.D. Princeton 1985. The determination of municipal budgets: Four empirical essays.

Bong Young Hong, Ph.D. Georgia State 1985. Multiobjective bank balance management under conditions of uncertainty.

Eero Kasanen, D.B.A. Harvard 1986. Capital budgeting and the control of business unit growth.

Naveen Khanna, Ph.D. Northwestern 1986. Optimal takeover and defensive strategies in markets for corporate acquisitions.

Stuart K. Landon, Ph.D. Queen's 1985. Three essays on adjustment.

Vincenzo La Via, Ph.D. California-Los Angeles 1985. Exchange rates and monetary regimes: Relearning the lessons of the 1920s.

Emily C. Lawrance, Ph.D. Yale 1986. The savings behavior of rich and poor: A study of time preference and liquidity constraints.

Jin Soon Lee, Ph.D. Wisconsin-Madison 1986. Taxation, bequest, and lifetime portfolio selection.

Andrew B. Lyon, Ph.D. Princeton 1986. How do stock prices change when corporate tax laws change? An empirical analysis of the effect of tax laws on the value of the firm.

James B. Mackie, Ph.D. Virginia 1986. Two essays on capital taxation.

Jeffrey K. Mackie-Mason, Ph.D. MIT 1986. Dynamic uncertainty, taxes, and risky investment decisions: Theory and application.

Guritno Mangkusubroto, Ph.D. Colorado-Boulder 1986. Tax incidence in a developing country: The case of Indonesia.

Milton H. Marquis, Ph.D. Indiana-Bloomington 1985. Uncertainty in money demand.

Mark J. Mazur, Ph.D. Stanford 1986. Three essays on income tax progressivity.

Richard H. Megna, Ph.D. Wisconsin-Milwaukee 1986. Estimation of the demand for local public education under a nonlinear budget constraint.

Daniel C. Messerschmidt, Ph.D. Iowa State 1984. Hedging mortgage risk with T-bond futures and options: A case study.

Masoud Moghaddam, Ph.D. Iowa State 1984. Two essays on money supply growth, inflation, and interest rates.

Abdol H. Moini, Ph.D. Alabama 1986. The usefulness of credit aggregates as intermediate monetary policy targets.

Leonard I. Nakamura, Ph.D. Princeton 1986. Financial intermediaries and real activity: Theory and evidence.

Vasco M. d.P. Orey, Ph.D. Illinois-Urbana 1986. Monetary games in interdependent economies.

Zulkifly B. Osman, Ph.D. Colorado-Boulder 1985. Towards an aggregate demand for money model under uncertainty.

Chwee-Huay Ow, Ph.D. Johns Hopkins 1986. The currency board monetary system: The case of Singapore and Hong Kong.

Marco Pagano, Ph.D. MIT 1986. Market size and asset liquidity in stock exchange economies.

Leonidas S. Pappas, Ph.D. SUNY-Binghamton 1986. Optimal taxation of alcohol and tobacco in the presence of externalities: An econometric analysis.

Pier L, Parcu, Ph.D. California-Los Angeles 1985. Financial instability and monetary institutions: Essays in monetary theory.

Young Joe Park, Ph.D. Pittsburgh 1986. Optimal commodity tax: Korean electricity case.

ROWENA A. PECCHENINO, Ph.D. Wisconsin-Madison 1985. Organization, coordination, and regulation: Three essays on the theory of banking.

BRUCE D. PHELPS, Ph.D. Yale 1985. A finance approach to convertible money regimes: A new interpretation of the greenback era.

GEORGE A. PLESKO, Ph.D. Wisconsin-Madison 1985. Environmental regulations, capital formation, and the economic performance of U.S. manufacturing industries.

GEOFFREY POITRAS, Ph.D. Columbia 1986. A study of gold futures price spreads.

DAVID G. RABOY, Ph.D. George Washington 1986. Capital recovery and economic efficiency: The case of imperfect loss offset.

S. SAJJADUR RAHMAN, Ph.D. Carleton 1985. The demand for money in developing economy: Evidence for six Asian countries.

ANDREW ROSE, Ph.D. MIT 1986. The autoregressivity paradox in macroeconomics.

ERIC S. ROSENGREN, Ph.D. Wisconsin-Madison 1986. Tax effects in the corporate and government bond markets.

STUART S. ROSENTHAL, Ph.D. Wisconsin-Madison 1986. Housing tax policy, residence times, and the cost of moving.

GEORGE A. SELGIN, Ph.D. New York 1986. The theory of free banking: A study of the supply of money under competitive note issue.

BRIAN SHEA, Ph.D. Maryland 1986. Stabilization and profitability of sterilized intervention in foreign exchange markets.

ANDREI SHLEIFER, Ph.D. MIT 1986. The business cycle and the stock market.

ROBERTA G. STEINMAN, Ph.D. Colorado-Boulder 1985. Financial innovations, the increased vulnerability of the liability structure, and the implications for monetary policy.

PAUL M. TAUBE, Ph.D. SUNY-Albany 1986. Temporary equilibrium in a private banking economy with credit rationing.

MILLICENT M. TAYLOR, Ph.D. Tennessee-Knoxville 1985. The monetary theories of Milton Friedman and J. M. Keynes.

CHRISTOPHER M. TOWE, Ph.D. Western Ontario 1985. Monetary versus fiscal policy: Three essays.

LENOS G. TRIGEORGIS, D.B.A. Harvard 1986. Valuing real investment opportunities: An options approach to strategic capital budgeting.

BRUCE N. TRUE, Ph.D. California-Los Angeles 1985. An examination of the monetary hypothesis of the depression and interest rates.

JOHN B. VAN HUYCK, Ph.D. Brown 1986. Seigniorage, sovereign, debt and reputation.

WILLIAM VAN LEAR, Ph.D. Pittsburgh 1986. Financial institutional control of nonfinancial institutions in the contemporary United States.

JOHN M. VEITCH, Ph.D. Northwestern 1985. Bank loan commitments, monetary policy, and real activity.

ROBERT W. VISHNY, Ph.D. MIT 1985. Informational aspects of securities markets.

MARK E. WOHAR, Ph.D. Illinois-Urbana 1985. Alternative modes of deficit financing, money growth, inflation, and interest rates.

ANDREW M. WOLFE, Ph.D. Wisconsin-Madison 1985. A strategic analysis of export performance requirements and investment incentives.

YONG JOON YOON, Ph.D. Northwestern 1986. Dynamic models for vintage capital goods and inventories.

JON R. ZUCKER, Ph.D. Stanford 1986. A theoretical and empirical investigation of asymmetry in corporate income taxation.

International Economics

NISHA AGRAWAL, Ph.D. Virginia 1986. Distributional effects of international debt: A general equilibrium model of the Phillipines.

STEVEN AMBLER, Ph.D. Stanford 1985. Announcement effects of fiscal policy in open economies.

SALIM BACHOU, Ph.D. Notre Dame 1986. Trade relations between the European Economic Community and its associates in the framework of LOME.

RICHARD BALDWIN, Ph.D. MIT 1986. Three essays in international trade and finance.

NANDI BANANI, Ph.D. New York 1986. Unemployment, growth, and structural change in a small open economy with rigid real wage.

BENOIT BLAREL, Ph.D. Wisconsin-Madison 1985. International trade theory and protection of the environment.

EDUARDO BORENSZTEIN, Ph.D. MIT 1986. Essays on fiscal policy, foreign debt, growth, and exchange rates.

IAN BORSOOK, Ph.D. Columbia 1985. Human capital, trade, and migration.

GUY A. BRIDGEMAN, Ph.D. Western Ontario 1985. International trade with quality differentiated goods.

NADEEM A. BURNEY, Ph.D. Johns Hopkins 1986. International capital movements and welfare in the presence of external economies of scale.

SANG K. BYUN, Ph.D. Columbia 1985. A comparison of shadow factor prices under different rates of tariff and its application in the welfare analysis of international factor mobility.

PUNAM CHUHAN, Ph.D. Georgetown 1986. Three essays on exchange rate policy.

BENEDICT J. CLEMENTS, Ph.D. Notre Dame 1986. Foreign trade structures, employment, income distribution, and the balance of payments: The case of Brazil.

CHARLES H. DALLARA, Ph.D. Tufts 1986. The role of the International Monetary Fund in the balance of payments adjustment process, 1974–76.

Paul R. DeLancey, Ph.D. Wisconsin-Madison 1986. Official reserve management and forecasts of official reserves.

Graciana Del Castillo, Ph.D. Columbia 1986. Balance of payments analysis: An econometric test of Uruguay.

Joaquim Detoledo, Ph.D. MIT 1986. Three essays on macroeconomic policies in the open economy.

Alex Donker, Ph.D. Florida State 1986. Monetary policy and the balance of payments: The case of Ghana, 1961–81.

John J. Fallon, Ph.D. Western Ontario 1986. System-wide estimation of trade elasticities for Canada and the United States: A comparison of the translog and differential approaches.

Mohsen Fardmanesh, Ph.D. Yale 1985. Terms of trade shocks and structural adjustment in a small open economy.

Karnit Flug, Ph.D. Columbia 1986. Government policies in a general equilibrium model of international trade and human capital.

Sumantra Ghoshal, D.B.A. Harvard 1986. The innovative multinational: A differentiated network of organizational roles and management processes.

Junichi Goto, Ph.D. Yale 1986. A general equilibrium analysis of international trade under imperfect competition in both product and labor markets: Theory and evidence from the automobile trade.

Vittorio U. Grilli, Ph.D. Rochester 1986. Studies in speculative attacks and investment banking contracts.

Frank Gunter, Ph.D. Johns Hopkins 1986. U.S. bank foreign lending.

Pertti Haaparanta, Ph.D. Yale 1986. External disturbances, exports, and wage indexation.

Frances R. Hammond, Ph.D. California-Los Angeles 1986. Industry wage response to product demand shifts.

Catharine B. Hill, Ph.D. Yale 1985. The effects of different trade policies in open economy macroeconomic models.

Chun-Tien Hu, Ph.D. Princeton 1986. Policy coordination and exchange rate policy in a world of flexible exchange rates.

Hyung-Doh Hur, Ph.D. Indiana-Bloomington 1985. Incremental efficiency and the impact of "news" on foreign exchange markets.

Yiu K. Ip, Ph.D. South Carolina 1985. Tests for international capital market integration.

Ioannis Kallianiotis, Ph.D. City-New York 1985. Exchange rates and rational expectations.

Roberto Khalil-Jalil, Ph.D. Wisconsin-Madison 1985. The determinants of Mexico's manufacturing exports (1950–83).

Mohiuddin M. Khan, Ph.D. Simon Fraser 1986. Active portfolio management and the gains from international portfolio diversification.

Shaheen R. Khan, Ph.D. Columbia 1986. Protection to large scale manufacturing industries in Pakistan.

Yooman Kim, Ph.D. Simon Fraser 1986. Theory of internationally diversified production under uncertainty: The effects of exchange rate fluctuations on growth of direct foreign investment, return, stability, and risk.

Paolo Kind, Ph.D. Columbia 1986. On models of money and perfect foresight and exchange rate dynamics.

Guillermo R. Le Fort-Varela, Ph.D. California-Los Angeles 1985. The real exchange rate and capital inflows: The case of the southern cone countries.

Robert E. Lippens, Ph.D. Queen's 1985. Multimaturity tests in international financial markets of efficient market hypotheses: A comparative study of Japan and Canada.

Bih-Jane Liu, Ph.D. Wisconsin-Madison 1985. Trade and tax policies in distorted small economies.

Judith A. McDonald, Ph.D. Princeton 1986. Factors influencing countries' decisions to use credit from the International Monetary Fund.

John Malindretos, Ph.D. Rutgers 1986. A theoretical and empirical comparison of the two theories of international finance: The case of the Federal Republic of Germany.

Michael H. Moffett, Ph.D. Colorado-Boulder 1985. The international interbank market: Credit rationing and bank-tiering under imperfect information.

Hillar Neumann, Ph.D. Rutgers 1986. Savings, capital formation and real exchange rates: An application of capital theory to balance of payments analysis and exchange adjustment.

Yasuo Nishiyama, Ph.D. California-Berkeley 1986. Exchange rates: Forward and backward linkages on U.S. agriculture.

Sean Nolan, Ph.D. Yale 1986. The multinational enterprise and the international division of labor: A theoretical analysis.

J. Marcus Noland, Ph.D. Johns Hopkins 1986. The determinants of comparative advantage: Empirical analysis in a multicountry time-series cross-section framework.

Patricia A. O'Brien, D.B.A. Harvard 1986. Coordinating market forces: The anatomy of investment decisions in the Japanese steel industry.

Sang B. Oh, Ph.D. Columbia 1985. Transfer of technology: Econometric analysis of Korea.

Sule Özler, Ph.D. Stanford 1985. Rescheduling of sovereign government bank debt.

Jeffrey L. Pliskin, Ph.D. Michigan 1985. Two essays on international capital flows.

James D. Reitzes, Ph.D. Wisconsin-Madison 1986. Essays on the welfare and policy implications of endogenous capital flows in two distinct sectors.

Barbara L. Rollinson, Ph.D. Duke 1986. Multiplant economies in multinational enterprises: Implications for taxation and commercial policy.

JEFFREY ROSENSWEIG, Ph.D. MIT 1985. Empirical essays on the dollar area.

L. FERNANDO RUIZ-MIER, Ph.D. Purdue 1985. Trade, technology transfer, and multinational enterprises in monopolistically competitive models.

MOHAMMED Z. SALEH, Ph.D. Colorado-Boulder 1985. A structural model of Malaysian balance of payments: 1963–82.

HAIDEH SALEHI-ESFAHANI, Ph.D. Pennsylvania 1985. "Dutch Disease": The perverse effects of a resource boom in a small open economy.

FERNANDO SANTOS, Ph.D. South Carolina 1985. Three essays on Portuguese monetary economics.

AUREL SCHUBERT, Ph.D. South Carolina 1985. The credit-Anstalt crisis of 1931: A financial crisis revisited.

MIGUEL SEBASTIAN, Ph.D. Minnesota 1985. Fixed exchange rates and noncooperative monetary policies.

MANUEL R. SEBASTIAO, Ph.D. Columbia 1986. Economics of scale, pricing policies, and international trade.

SELIG SECHZER, Ph.D. Pennsylvania 1986. Essays in factor mobility.

LUIS SERVEN, Ph.D. MIT 1986. Essays on exchange rates and stabilization policy.

RAMA SETH, Ph.D. Columbia 1986. Optimal foreign borrowing: The case of Costa Rica.

CLINTON R. SHIELLS, Ph.D. Michigan 1985. A disaggregated empirical analysis of U.S. import demand, 1962–81.

OZ SHY, Ph.D. Minnesota 1986. International trade restrictions that improve world welfare.

ILTAE SON, Ph.D. Northwestern 1985. The choice of an optimal exchange rate system for a small country.

ROBERT W. STAIGER, Ph.D. Michigan 1985. Essays in international trade.

ANDREW D. STEER, Ph.D. Pennsylvania 1985. Exchange rate expectations and the market for foreign exchange: A study of the sterling-dollar rate.

GERARD L. STOCKHAUSEN, Ph.D. Michigan 1985. The effects of a quota threat on the exporting country.

RANJIT S. TEJA, Ph.D. Columbia 1986. International capital controls: Essays on the theory of floating dual exchange rates.

HENDRIKUS TRINES, Ph.D. Pittsburgh 1985. Exchange rates, output, and prices in an open economy.

CARLOS M. URZUA, Ph.D. Wisconsin-Madison 1986. Foreign debt and the less developed countries: An empirical-theoretical examination of the failure of a market.

JAMES E. VESTAL, Ph.D. Michigan 1985. An analysis of the determinants and commodity trade effects of Japanese technology trade.

RANDALL M. WIGLE, Ph.D. Western Ontario 1986. General equilibrium analysis of the effects of multilateral and Canadian liberalization.

RONALD N. WOODS, Ph.D. Florida State 1986. The impact of the economist upon tariff policy in Australia: A half-century of tariff debate.

FARAMARZ YAZDANI, Ph.D. Stanford 1986. Japanese trading companies: Their role in foreign trade invoicing.

STEPHEN D. YOUNGER, Ph.D. Stanford 1986. Inflation-indexed exchange rates and macroeconomic stability in less developed countries.

ANDREW B. ZIMMERMAN, Ph.D. California-Berkeley 1985. The relationship between economic indicators and country risk.

Business Administration; including Business Finance and Investment, Insurance Marketing, and Accounting

CHARLES W. ABDALLA, Ph.D. Michigan State 1985. Problems in interindustry coordination and system-wide productivity innovation: An institutional analysis of barriers to implementing modular shipping containers for dry groceries.

JOHN C. ANDREWS, Ph.D. South Carolina 1985. Tests of the elaboration likelihood model involving persuasive marketing communication.

MARY R. BEJAN, Ph.D. California-Berkeley 1985. The differential impact of noisy and noiseless public information on social welfare.

JULIA BRAZELTON, Ph.D. South Carolina 1985. Tax implications: The implications of a consumptive tax base on the individual taxpayer.

CHARLES B. CADSBY, Ph.D. MIT 1985. Essays on adjustment of equilibrium in financial asset markets.

ROBERTA L. COOK, Ph.D. Michigan State 1985. The Mexican dry grocery subsector: Strategies supporting the establishment of voluntary food chains.

RITA DRIEGHE, Ph.D. Northwestern 1986. Boards of directors and corporate governence: The case of Greenmail.

SRVINIVAS DURVASULA, Ph.D. South Carolina 1986. Modeling the multiple store location allocation problem: Alternate approaches.

FRED O. EDE, Ph.D. South Carolina 1985. The effectiveness of persuasion techniques as negotiation strategies in buyer-seller interaction.

STEVEN L. FLINT, Ph.D. Stanford 1986. Q-ratios and property-liability insurance.

ANTONIO R. FRANCES, Ph.D. California 1986. The dynamics of economic adjustment in business firms: The response of Venezuelan exporters to the exchange control of 1983 as a test case.

CATHY GOODWIN, Ph.D. California-Berkeley 1986. Using role relationship dimensions to predict forms of complaint behavior of service consumers.

JAVAD GORJIDOOZ, Ph.D. Indiana-Bloomington 1985. The stock portfolio performance of property-liability insurance stock companies: Using the *K*-factor model of arbitrage pricing theory.

MINHI HAHN, Ph.D. Northwestern 1985. Advertising models for new products: New models and their validation.

PETER R. HAMILTON, Ph.D. Minnesota 1985. The applied welfare economics of university registration: A study in multi-object auctions.

CHEONG KON HAN, Ph.D. Northwestern 1985. Diversification strategies of multiproduct firms.

DAWN A. HARRIS, Ph.D. Northwestern 1986. Executive succession, reputation, and wage revision.

J. RICHARD HARRISON, Ph.D. Stanford 1986. The committee structure of corporate boards of directors.

MEIR KARLINSKY, Ph.D. California-Berkeley 1986. The case of decisions under risk.

RACHID LARAQUI, Ph.D. California-Berkeley 1985. Essays on futures markets and options.

MARJO-RIITTA LEHTISALO, Ph.D. Northwestern 1986. The influence of price on product evaluation: An information-processing perspective.

KIAN-GUAN LIM, Ph.D. Stanford 1986. Essays in financial economics.

VICTOR S. LIMLINGAN, D.B.A. Harvard 1986. The overseas Chinese in ASEAN: Business strategies and management practices.

JOSEPH YUET-CHANG LIN, Ph.D. Minnesota 1986. Three essays on the economics of marketing.

JOSEPH MAGLIOLO III, Ph.D. Stanford 1986. Oil and gas stock prices and reserve recognition accounting.

AJAY K. MANRAI, Ph.D. Northwestern 1986. Similarity perceptions and choice.

NAHUM MELUMAND, Ph.D. California-Berkeley 1985. Advances in the theory of managerial compensation.

JOAN MEYERS-LEVY, Ph.D. Northwestern 1986. Gender differences in information processing: A selectivity interpretation.

BENJAMIN M. OVIATT, Ph.D. South Carolina 1985. The influence of industry structure and business strategy on the relationship between risk and return.

SHEILA M. PUFFER, Ph.D. California-Berkeley 1985. Altruism in organizations: An examination of extra-role behavior in a retail store context.

RAMACHANDRAN RAMANAN, Ph.D. Northwestern 1986. Managerial incentives, accounting for interest costs, and capital investment decisions of the firm.

GONZALO A. RUBIO, Ph.D. California-Berkeley 1985. Asset pricing and equity rights issues: The case of the Spanish capital market.

MARK S. RZEPCZYNSKI, Ph.D. Brown 1986. A study of speculative behavior for primary government securities through a market-based model of financial futures trading.

YOSHIKIYO SAKAI, Ph.D. Minnesota 1986. A reconsideration of asset-pricing models.

JOHN P. SNYDER, Ph.D. South Carolina 1985. Analytical review: An empirical study.

AHMET I. SOYLEMEZOGLU, Ph.D. Catholic 1986. Multiperiod portfolio strategies using quadratic cost function to account for adjustment costs.

DOUGLAS M. STAYMAN, Ph.D. California-Berkeley 1985. Emotional response to advertising: Beyond attitude toward the ad.

LARRY E. SWEENEY, Ph.D. West Virginia 1985. An empirical evaluation of the debt capacity motive for conglomerate merger.

ROBERT J. SWEENEY, Ph.D. South Carolina 1985. Arbitrage pricing theory: An examination of the skewness effect.

NIRVALUR VIJAYARAGHAVAN, Ph.D. Stanford 1986. Essays on capital structure.

ROBERT VILLARREAL, Ph.D. MIT 1986. Investment and financing interactions at the firm level: An econometric simultaneous-equation approach.

MARTCIA E. WADE, Ph.D. California-Berkeley 1986. An equilibrium model of the market for physician services.

WILLARD M. G. WAGNER, Ph.D. California-Berkeley 1985. Contributions to collective goods model and experiment.

Industrial Organization and Public Policy; including Economics of Technological Change, and Industry Studies

CHARLES R. ADELBERG, Ph.D. Minnesota 1985. The market place for *R&D* outcomes: Its function and significance in the innovation.

JUSTO AGUILAR, Ph.D. Connecticut 1985. Technological change and governmental policies in the industrial development of Costa Rica.

MCKINLEY ALEXANDER, JR., Ph.D. Illinois-Urbana 1986. The nature of competition in highway freight transportation.

MARIE C. ANCHORDOGUY, Ph.D. California-Berkeley 1986. Japanese industrial policy towards the computer industry.

MARK E. BAGNOLI, Ph.D. Princeton 1985. Equilibrium in product markets containing imperfectly informed consumers.

MITCHELL BERLIN, Ph.D. Pennsylvania 1986. Screening, commitment, and the choice between long- and short-term contracts.

PHILIP A. BEUTEL, Ph.D. Duke 1986. Cable television franchising: A model of bidding firm behavior and auction resolution.

Swati Bhatt, Ph.D. Princeton 1986. Nonprice competition in oligopolistic industries.

Marsha A. Blumenthal, Ph.D. Minnesota 1985. Auctions with constrained information: Blind bidding for motion pictures.

Scott J. Callan, Ph.D. Texas A&M 1985. An empirical examination of alternative productivity growth measures, with an application to the steam-generated fossil-fueled electric utility industry, 1951–78.

Duncan J. Cameron, Ph.D. California-Los Angeles 1985. The Chicago Board of Trade case.

Pamela Cameron, Ph.D. Oklahoma 1985. Testing the validity of using the Posner framework for detecting collusive price behavior using economic evidence.

Charles Capone, Ph.D. Virginia 1985. Technology of energy production in cement manufacture.

Nilufer Cagatay, Ph.D. Stanford 1986. The interindustry structure of wages and markups in Turkish manufacturing.

Siriwan Chutikamoltham, D.B.A. Harvard 1986. A price analysis of the Thai corn commodity.

Paul A. Cleveland, Ph.D. Texas A&M 1985. Market performance under deregulation: An industry study of air transportation.

Stephen C. Cooke, Ph.D. Michigan State 1985. The theory of consumers' surplus and economic rent: An application for measuring the benefits of the mechanical cucumber harvester.

Kevin M. Currier, Ph.D. SUNY-Albany 1986. Optimal pricing and regulation under incomplete information.

Fekru Debebe, Ph.D. Colorado State 1985. Nonlinear estimation of variability for rail rate regulation under the Staggers Act.

Soctt M. Dennis, Ph.D. Northwestern 1985. The effect of end-to-end railroad mergers on economic welfare.

Thomas W. Dorsey, Ph.D. Wisconsin-Madison 1985. Monopolistic competition in high technology industry.

Sadok Driss, Ph.D. Illinois 1986. Evaluation research and agricultural development: A Tunisian case study.

Hossein Eftekhari, Ph.D. Nebraska-Lincoln 1985. The Averch-Johnson thesis and the fair rate of return: An exercise in multi-objective programming.

Steven L. Estomin, Ph.D. Maryland 1986. The determinants of profitability and merger premiums in congolmerate mergers.

Reza Fadei-Tehrani, Ph.D. Nebraska-Lincoln 1986. A study of the optimum location of manufacturing activities in the air compressor industry.

Alan E. Finder, Ph.D. Indiana-Bloomington 1985. Empirical tests of cost subadditivity in the investor-owned electric utility industry.

Mary E. Fitzpatrick, Ph.D. Northwestern 1986. The process of regulation: Theory and evidence.

Glenn C. Fox, Ph.D. Minnesota 1985. Optimal public investment in U.S. agricultural research: A study in the management of technical change.

Murray Z. Frank, Ph.D. Queen's 1986. Essays on industrial behavior under uncertainty.

Steven R. Galginaitis, Ph.D. SUNY-Albany 1986. Capital-energy complementarity, economies of scale, and technological change in U.S. manufacturing: A regional analysis.

Sarah J. Goodfriend, Ph.D. North Carolina-Chapel Hill 1985. Estimation of a Q-ratio function for regulated electric utilities: A test of the Stigler-Peltzman hypothesis of regulatory behavior.

James Hagerman, Ph.D. Northwestern 1985. Issues in regulation by price adjustment.

John W. Handy III, Ph.D. Georgia State 1985. A macro analysis of growth and change in the black-owned business sector.

Richard S. Harris, Ph.D. California-Los Angeles 1985. An economic analysis of first-refusal and related preemptive rights.

John D. Henderson, Ph.D. Western Ontario 1986. Economic performance in the Canadian tomato-processing industry.

M. Anisul Islan, Ph.D. Alberta 1985. Monopoly and competition in the real estate brokerage industry: Theory and evidence.

Si-Young Jang, Ph.D. Texas A&M 1986. Oligopolistic market behavior under deregulation.

Gary H. Jefferson, Ph.D. Yale 1986. The sources of productivity growth: A dynamic model of short- and long-run productivity change.

William E. Johnson, Ph.D. South Carolina 1985. Strategic groups and market rivalry.

Susan K. Jones, Ph.D. Wisconsin-Madison 1985. Import competition and wages in U.S. manufacturing industries: Theoretical perspectives and empirical results using a simultaneous system.

Kyungwoo Kang, Ph.D. Pennsylvania 1985. Impacts and implications of the motor carrier deregulation.

John A. Karikari, Ph.D. Queen's 1985. Tariff protection in oligopoly: A theoretical and empirical analysis.

Tadahiko Kawai, Ph.D. California-Berkeley 1986. An analysis of business-government relations in Japan: The case of the Federation of Economic Organizations.

Woo Bong Kim, Ph.D. California-Berkeley 1986. The efficiency of a group of machines served by a single server.

Alison Kirby, Ph.D. Stanford 1986. Trade associations as information exchange mechanisms.

Nan S. Langowitz, D.B.A. Harvard 1986. An exploration of production problems in the initial commercial manufacture of new products.

Necia Lanius, Ph.D. South Carolina 1985. Competition and pricing in the hospital industry.

DANIEL F. LEACH, JR., Ph.D. California-Los Angeles 1985. The social cost of industrial deconcentration.

CHOON SEI LEE, Ph.D. Texas A&M 1986. The impacts of price discrimination on the differentiated industry.

PAUL R. LEFRANCOIS, Ph.D. West Virginia 1985. Average versus marginal cost pricing for water service.

MARK E. LEVONIAN, Ph.D. MIT 1986. Steel minimills and the spatial characteristics of steel demand.

J. NELLIE LIANG, Ph.D. Maryland 1986. An empirical conjectural variation model of oligopoly.

YEO LIN, Ph.D. Northwestern 1986. Technological change: A microeconomic study of the consumer electronics industry in Taiwan.

DANIEL W. MCCOLLUM, Ph.D. Wisconsin-Madison 1986. A travel cost methodology: Travel time in the traditional travel cost model, development of a probabilistic travel cost model, and comparisons with a simulated market.

ELLEN B. MAGENHEIN, Ph.D. Maryland 1986. The gains to acquisition, acquisition premia, and the theory of the firm: A theoretical and empirical analysis.

JEFF D. MAKHOLM, Ph.D. Wisconsin-Madison 1986. Sources of total factor productivity in the electric utility industry.

FARIED W. MANSOER, Ph.D. Colorado 1986. Electricity pricing and investment under government policy constraints: The case of the Java-Indonesia interconnected supply system.

CRISTINA MAZON, Ph.D. Minnesota 1985. The impact of government policy on the U.S. steel and cigarette industries.

GEORGE W. MECHLING, Ph.D. Nebraska-Lincoln 1985. A systems concept approach to the explanation and prediction of technological change: An econometric study.

LORETTA J. MESTER, Ph.D. Princeton 1985. Three essays on industrial organization: An empirical and theoretical analysis of the structure and behavior of savings and loans.

JOHN E. MORGAN, JR., Ph.D. Georgia 1986. A computerized economic and statistical analysis of the Alabama asphalt market.

JOHN R. MORRIS, Ph.D. Washington 1985. Intellectual property: Creating, pricing, copying.

JULIANNE B. NELSON, Ph.D. Princeton 1986. Contract enforceability and Ramsey optimal cost recovery in a natural monopoly.

THOMAS R. OVERSTREET, JR., Ph.D. Vanderbilt 1986. An economic analysis of resale price maintenance.

NICHOLAS C. PAPANDREOU, Ph.D. Princeton 1986. Price controls and industry structure in Greece: A theoretical and empirical analysis.

PHILIP G. PARDEY, Ph.D. Minnesota 1986. Public sector production of agricultural knowledge.

JUNG KEUN PARK, Ph.D. Minnesota 1986. Technological change in the Korean rice economy: Sources, direction, and impact.

GREGORY L. POWELL, Ph.D. Northwestern 1986. Competition in the U.S. petroleum refining industry: Empirical evidence and policy implications.

IRENE POWELL, Ph.D. Wisconsin-Madison 1985. Effect of reducing manufacturing concentration on the distribution of income: A simulation.

RUSSELL W. RADFORD, D.B.A. Harvard 1986. Patterns of performance: The development of technological capability in a process-oriented company.

JOHN J. REARDEN, Ph.D. Illinois-Urbana 1985. A theoretical and empirical analysis of public and private antitrust enforcement.

BRIAN E. ROBERTS, Ph.D. Washington-St. Louis 1986. Asset values and the political process.

NANCY ROSE, Ph.D. MIT 1985. The incidence of rents in the regulated trucking industry.

RAYMOND D. SAUER, Ph.D. Washington 1985. Credible information? The impact of the FTC's advertising substantiation program.

LAURENCE C. SCHUMANN, Ph.D. Virginia 1986. Mergers and shareholder returns.

GARY L. SHOESMITH, Ph.D. Southern Methodist 1986. The cost-minimizing industry structure for petroleum refining: An application of contestable market theory and multiproduct cost functions.

MARION SILLAH, Ph.D. South Carolina 1986. A theoretical model for the systematic analysis of an office space market.

JEFFREY W. SMITH, Ph.D. Northwestern 1986. A theoretical analysis of the supply of housing.

MARGARET D. SOMMERLATT, Ph.D. West Virginia 1986. The traffic sensitivity of nontraffic sensitive telephone exchange plant.

PHILIP SPILBERG, Ph.D. Maryland 1985. Mergers and aggregate concentration.

STEVEN J. TAFF, Ph.D. Wisconsin-Madison 1985. Evaluating product diversity in the U.S. pesticide industry.

RICHARD P. THEROUX, Ph.D. Purdue 1985. The effects of Ohio's 1983 driving-while-intoxicated law: A theoretical and empirical investigation.

PING SHU TSO, Ph.D. Washington-St. Louis 1986. Theories of investment and financing.

DAH-NEIN TZANG, Ph.D. Illinois-Urbana 1986. An analysis of financial contracting between multinational companies and developing countries for oil extraction.

DOUGLAS J. WAKEMAN, Ph.D. North Carolina-Chapel Hill 1986. An empirical investigation of abnormal returns in corporate mergers.

SATOSHI WATANABE, Ph.D. Princeton 1985. Public enterprises under imperfect competition.

Agriculture and Natural Resources

BEL H. ABDELKAFI, Ph.D. Missouri-Columbia 1985. Farm level impacts of a conservation easement program.

MOHAMED T. ALSAFFY, Ph.D. Washington State 1985. Demand for wheat and rice in selected Arab countries: The current situation and projections to year 2000.

ROBERT A. ANDROKOVICH, Ph.D. Western Ontario 1985. The impact of risk on contract structure in a principal-agent model: An application to the Alberta sugar beet industry.

SYED M. ARIF, Ph.D. Illinois 1986. Economics of decision making in agricultural pest management.

BRIAN P. BAKER, Ph.D. Cornell 1985. Protection of groundwater from agricultural pollution: Institutions and incentives.

MAMADOU BAMBA, Ph.D. Illinois 1985. Stabilizing the cocoa market: A quarterly econometric application using optimal control theory.

JEFFREY R. BEAULIEU, Ph.D. Iowa State 1984. The projected impact of inland waterway user charges on grain flows and transport costs in 1990.

DANIEL J. BERNARDO, Ph.D. Washington State 1985. Optimal irrigation management under conditions of limited water supply.

ABHA BHARGAVA, Ph.D. Alberta 1986. The effects of public policy on alberta's oil and gas industry.

KEVIN J. BOYLE, Ph.D. Wisconsin-Madison 1985. Essays on the valuation of nonmarket resources: Conceptual issues and empirical case studies.

ROBERT F. BROOKER, Ph.D. North Carolina State 1985. Multiple component pricing and the composition of milk: An econometric model.

TERRY N. BROWNSCHIDLE, Ph.D. Minnesota 1985. An evaluation of the economic incentives of water use charges for producers under uncertainty.

RICHARD T. CARSON, JR., Ph.D. California-Berkeley 1985. Three essays on contingent valuation.

DAVID E. CLARK, Ph.D. SUNY-Binghamton 1986. Evaluating the recreational benefits of environmental improvement using compensating wage differentials.

KATHLEEN L. COLE, Ph.D. Cornell 1986. An integrated electric utility planning model: An illustration of the costs of financing acid deposition controls.

IAN L. DALZIELL, Ph.D. Michigan State 1985. Sources of agricultural instability.

EARL L. DEAL, JR., Ph.D. North Carolina State 1986. Competitiveness of alternative forest products manufacturing facilities in purchasing timber inputs.

MASSOUD S. M. DENBALY, Ph.D. Iowa State 1984. U.S. monetary policy and the exchange rate: Effects on the world coarse grain market.

BERNADETTE DIA, Ph.D. Illinois 1986. Default and liquidity in the management of the BNDA lending program in Ivory Coast.

MICHAEL R. DICKS, Ph.D. Missouri-Columbia 1985. Aggregate economic impacts of a conservation easement program for the Corn Belt.

TERRY M. DINAN, Ph.D. Iowa State 1984. Obtaining the optimal fuel-conserving investment mix: A linear programming-hedonic technique approach.

BRIAN K. EDWARDS, Ph.D. California-San Diego 1984. Essays in the economic analysis of the multispecies fishery.

ERIC E. ELDER, Ph.D. Washington State 1985. Economic impacts of irrigation development in Washington.

JEFFREY E. ENGLIN, Ph.D. Washington 1986. Backcountry hiking and optimal forest management.

BRUCE E. FLORY, Ph.D. Wisconsin-Madison 1986. The impact of farm financial stress on investment in erosion control practices.

JAMES B. FRENCH, Ph.D. Missouri-Columbia 1985. Optimal disease management: A dairy herd application.

MURRAY E. FULTON, Ph.D. California-Berkeley 1985. Public enterprises in natural resource industries: An economic analysis.

WALTER H. GARDINER, Ph.D. Purdue 1986. The impact of alcohol fuel production on agricultural markets.

CAROL C. S. GILBERT, Ph.D. North Carolina-Chapel Hill 1985. Household adjustment and the measurement of benefits from environmental quality improvements.

JANE E. GLEASON, Ph.D. Illinois 1986. Decision making, liquidity, and risk under multiple-cropping conditions in Taiwan.

PAUL W. HEISEY, Ph.D. Wisconsin-Madison 1985. Employment and income in Botswana's arable agriculture.

STEPHEN W. HIEMSTRA, Ph.D. Michigan State 1985. Labor relations, technological and structural change in U.S. beef packing and retailing.

CHUNG-HUANG HUANG, Ph.D. Minnesota 1986. The recreation benefits of water quality improvement in selected lakes in Minnesota.

BENGT T. HYBERG III, Ph.D. North Carolina State 1986. The development and empirical testing of a utility maximization model for NIPF landowner forest management.

CARLOS A. IBANEZ-MEIER, Ph.D. Iowa State 1985. Economics of conservation tillage and soil productivity in north central Iowa: A control theory approach.

CRAIG E. JAGGER, Ph.D. Cornell 1986. Addressing excess capacity in wheat production: An economic analysis of long-range retirement programs.

REBECCA L. JOHNSON, Ph.D. Michigan State 1985. The effect of user fees on Great Lakes grain transportation.

CLIFTON T. JONES, Ph.D. Texas A&M 1985. Texas petroleum production and the windfall profit tax.

HARRY M. KAISER, Ph.D. Minnesota 1985. An analysis of farm commodity programs as risk management strategies for Minnesota corn and soybean producers.

GORDON M. KARP, Ph.D. SUNY-Albany 1986. Essays on the economic control of pollution.

MICHAEL S. KAYLEN, Ph.D. Purdue 1985. Times-series modeling of the live hog market: Forecasts and analysis of dynamics.

MARK A. KELLOGG, Ph.D. Washington 1985. Implicit valuation of flood hazard levels.

THOMAS A. KNAPP, Ph.D. Colorado 1985. Towards a synthesis of alternative amenity valuation techniques.

ANTHONY KRAUTMANN, Ph.D. Iowa 1985. The economics of nuclear power.

JAE-OK LEE, Ph.D. North Carolina State 1985. Japanese agricultural policies and import demand for wheat.

JAMES D. LEIBY, Ph.D. North Carolina State 1986. An analysis of farm size and growth: The case of southern dairy farms.

DONALD JIANN-TYNG LIU, Ph.D. Minnesota 1985. The willingness to pay of an individual fed cattle producer for a better forecasting technique: An evaluation in a dynamic economic setting with cash and futures marketing alternatives.

RAMATU MAHAMA, Ph.D. Iowa State 1985. A stochastic simulation of the impact of price insulation policies on world wheat market stability.

KEITH L. MENZIE, Ph.D. Purdue 1986. An analysis of factors affecting the organization of a cooperative feed manufacturing and distribution system.

JOAN K. MEYER, Ph.D. Cornell 1985. Effect of futures trading on the supply of lumber.

SCOTT R. MILLIMAN, Ph.D. Wisconsin-Madison 1985. Exploring the economics of resource rehabilitation within an investment framework: A case study of a fishery restoration process.

ROBERT J. MYERS, Ph.D. Minnesota 1986. Economic inefficiency under uncertainty and the effects of ideal risk markets in U.S. agriculture.

KATHLEEN A. MILLER, Ph.D. Washington 1985. The right to use vs. the right to sell: Water rights in the western United States.

JAMES W. MJELDE, Ph.D. Illinois 1985. Dynamic programming model of the corn production decision process with stochastic climate forecasts.

M. FAUZI MOHDJANI, Ph.D. Washington State 1985. An econometric analysis of supply and demand of fresh milk in pennisular Malaysia.

KEVIN C. MOORE, Ph.D. Iowa State 1985. A predictive econometric model of the U.S. farmland market: A rational expectations approach.

TAREK A. MOURSI, Ph.D. California-Berkeley 1986. Government intervention and the impact on agriculture: The case of Egypt.

DARYL NALL, Ph.D. Florida State 1986. Prospects for industrial cogeneration in Florida: The effects of regulatory policies.

CHARLES H. NELSON, Ph.D. Purdue 1986. Resource allocation and crop insurance.

SUSAN C. NUNN, Ph.D. Wisconsin-Madison 1986. The political economy of institutional innovation: Coalitions and strategy in the development of groundwater law.

ABDALLAH OMEZZINE, Ph.D. Missouri-Columbia 1985. Market analysis of the winter vegetable industry in Tunisia.

DAVID W. PARK, Ph.D. Cornell 1986. An appraisal of composite forecasting methods with applications to agricultural commodity price series.

BURTON W. PFLUEGER, Ph.D. Illinois 1985. Lender response to federal crop insurance: Their effects on farm business performance.

BASSAM I. RAMADAN, Ph.D. Cornell 1986. Food security in the Arab world: Its significance and means of attainment.

CLIFFORD V. ROSSI, Ph.D. Cornell 1986. An institutional and economic analysis of rural water supply infrastructure deterioration in New York State.

JEREMIAH B. I. RUGAMBISA, Ph.D. Illinois 1985. Structural change in the U.S. soybean and soybean products markets: A systematic varying coefficient simultaneous system approach.

KIMSEYINGA SAVADOGO, Ph.D. Purdue 1986. An analysis of the economic and sociodemoraphic determinants of household food consumption on Ouagadougou, Burkina, Faso.

MUSTAFA A. SAWANI, Ph.D. Missouri-Columbia 1985. Some effects of capital rationing and interest rates on farm organization and income.

PARVEEN P. SETIA, Ph.D. Illinois 1985. Simulating the adoption of soil conservation management systems under the conditions of uncertainty.

RADWAN A. SHABAN, Ph.D. Stanford 1985. Agricultural land tenancy with endogenous contracts: A theoretical and empirical investigation.

WILLIAM D. SHAW, JR., Ph.D. Colorado 1985. A utility-theoretic approach to estimating the demand for the benefit from recreational fishing: The impact of acid rain.

KOSTAS G. STAMOULIS, Ph.D. California-Berkeley 1986. The effects of monetary policy on U.S. agriculture: A fix-price flex-price approach.

HABIBAH SULEIMAN, Ph.D. Missouri-Columbia 1985. An econometric model of the Malaysian natural rubber industry: A phenological approach.

JEAN C. SUSSMAN, Ph.D. Minnesota 1985. A comparison of part-time and full-time agricultural operations in Dodge County, Minnesota.

MARTIN TURNER, Ph.D. Missouri-Columbia 1985. Measuring genetic improvement in soybean varieties and its influence on aggregate yields in the U.S. Corn Belt.

KEVIN D. WALKER, Ph.D. Missouri-Columbia 1985. Economic analysis of the effects of paratuberculosis (Johne's disease) in dairy herds.

PATRICK G. WELLE, Ph.D. Wisconsin-Madison 1986. Potential economic impacts of acid deposition: A contingent valuation study of Minnesota.

MICHAEL P. WELSH, Ph.D. Wisconsin-Madison 1986. Exploring the accuracy of the contingent valuation method: Comparisons with simulated markets.

MARCIA N. WHITE, Ph.D. Illinois 1986. Limited resource countries and agricultural development: A methodology used for the Caribbean.

Manpower, Labor, and Population, including Trade Unions and Collective Bargaining

LAURENCE BALL, Ph.D. MIT 1986. Essays on the labor market.

DEBRA A. BARBEZAT, Ph.D. Michigan 1985. Topics in academic labor markets.

MARY P. BECKMAN, Ph.D. Notre Dame 1986. Job rationalization, depersonalization and control: Clerical work and VDT's.

DALE L. BELMAN, Ph.D. Wisconsin-Madison 1986. The direct and indirect effects of industry structure on union membership and wages.

MARTI S. BOGART, Ph.D. Purdue 1985. Duration of unemployment spells ending in withdrawal: Theory and evidence.

FREDERICK C. BOLD, Ph.D. Washington 1985. Effort in human capital formation.

PAUL D. BOLDIN, Ph.D. Wisconsin-Madison 1985. Race, unemployment, and the wages of adult men: A longitudinal analysis.

I. BUSTILLO, Ph.D. American 1985. An assesssment of the productivity effect of the shift of women workers.

PAUL CARLIN, Ph.D. Pittsburgh 1985. The importance of bargaining phenomena in the economics of marriage and household decision making: Theoretical and empirical explorations.

RACHEL E. CONNELLY, Ph.D. Michigan 1985. Two essays on demographics and earnings.

BEN CRAIG, Ph.D. Stanford 1985. Modeling and estimating strikes as wars of attrition.

MASSOUMEH M. DIBA, Ph.D. Oklahoma 1985. The effect of interruption on earning and labor force participation of married women.

DENISE E. DIMON, Ph.D. Illinois-Urbana 1986. The interrelationship of immigration, trade, and employment.

MUHAMAD DJUHARI, Ph.D. George Washington 1986. The impact of governmental interventions on household behavior, with Indonesia as a case study.

TORBEN DREWES, Ph.D. Queen's 1985. Interregional wage transmission and unemployment disparities in Canada.

KATHLEEN M. GILLESPIE, Ph.D. Washington-St. Louis 1986. The effects of a copayment on disenrollment and utilization in a prepaid group practice.

LONNIE M. GOLDEN, Ph.D. Illinois-Urbana 1985. Did labor market forces cause the unemployment of the Great Depression?

ELIZABETH HILL, Ph.D. Maryland 1985. Postsecondary vocational education in Pennsylvania: In school and on the job.

KEITH HYLTON, Ph.D. MIT 1986. Essays on employment and real wages.

GAIL A. JENSEN, Ph.D. Minnesota 1986. Employer-sponsored insurance: A theoretical and empirical analysis.

FENG-FUH JIANG, Ph.D. SUNY-Binghamton 1986. Human capital investment as a determinant of life cycle labor supply.

MARK A. JOHNSON, Ph.D. Stanford 1985. The political economics of labor segmentation: The development of the California labor market, 1850–1900.

SUNG PIL KANG, Ph.D. South Carolina 1986. Implicit contracts on the job training and differences in age earnings profiles by industry.

J. KAPLAN, Ph.D. American 1985. The divide and conquer hypothesis: An analysis of internal labor markets and racial inequality of men in the United States.

TAKAO KATO, Ph.D. Queen's 1986. An economic analysis of Japanese and North American internal labor markets.

LAWRENCE F. KATZ, Ph.D. MIT 1986. Worker mobility and unemployment.

GENEVIEVE M. KENNEY, Ph.D. Michigan 1986. The security demand for children in Peninsular Malaysia.

SAM C. KOH, Ph.D. South Carolina 1985. The effects of ill health on wage rates and hours worked.

ROBERT J. LALONDE, Ph.D. Princeton 1985. Evaluating the econometric evaluations of training programs with experimental data.

ROBERT J. LESSNE, Ph.D. Pacific Western 1985. A study in compensation economics and ergonomics.

MARY R. MCCARTHY, Ph.D. New York 1985. An analysis of U.S. labor market flows and stocks.

JORGE O. MARISCAL, Ph.D. New York 1986. The economics of international migration: Mexican undocumented migrants in the United States.

KEIICHIRO MATSUSHITA, Ph.D. Michigan 1986. An economic analysis of age at first marriage.

RONALD B. MINCY, Ph.D. MIT 1986. Minimum wage noncompliance, enforcement, and establishment size.

JEAN M. MITCHELL, Ph.D. Vanderbilt 1986. The effect of arthritis on work behavior.

KOICHI NAKAJIMA, Ph.D. Vanderbilt 1986. Dynamic models of optimal rural family formation.

IGNACE NG-YIN-PING, Ph.D. Simon Fraser 1986. Transaction costs, labor market institutions, and strikes.

SOLEDAD NUNEZ, Ph.D. Minnesota 1986. Unemployment in the OECD countries: Estimating and modeling the effects of the proportion of population with education.

MAGDALENA RAPPL, Ph.D. South Carolina 1985. Age as a factor in the net earnings impact of CETA, and an analysis of the selectivity problem.

DAVID REIFFEN, Ph.D. California-Los Angeles 1985. A model of heterogeneous labor markets.

CURTIS J. SIMON, Ph.D. SUNY-Binghamton 1986. A portfolio theory of unemployment.

RICHARD D. STOCK, Ph.D. Colorado 1986. The intrametropolitan movement and metropolitan inmigration of residents by income and demographic characteristics and intrametropolitan changes in housing, employment, and income, 1950–80: An econometric analysis.

PATRICIA L. STUART, Ph.D. Colorado 1985. The short-run determinants of U.S. immigration: An econometric analysis.

EILEEN TRCZINSKI, Ph.D. Michigan 1985. Effect of uncertainty and risk on the allocation of time of married women.

WILLIAM J. WALSH, Ph.D. Indiana-Bloomington 1986. An institutional and economic analysis of the teacher collective bargaining act.

ROGER WEIKLE, Ph.D. South Carolina 1985. The determinants of local union leader attitudes and preferences for adjustment to technological change.

WILLIAM P. WELCH, Ph.D. Washington 1985. Wages, firm size, and the capital-labor ratio.

PAUL W. WILSON, Ph.D. Brown 1986. Income variation, the value of travel time, and implicit scheduling costs in staggered work-hours equilibria.

PHANINDRA V. WUNNAVA, Ph.D. SUNY-Binghamton 1986. The effect of unions on the level and slope of age-earnings profile in a life cycle framework: Evidence from panel data.

Welfare Programs; Consumer Economics; Urban and Regional Economics

PARVEZ AZIM, Ph.D. Texas A&M 1986. Impact of different pricing policies and types of competition on the optimum location of industry.

SUN IL BARK, Ph.D. Wisconsin-Madison 1985. The effect of net social wealth on the accumulation and distribution of nonhuman wealth.

NAJLA BATHISH, Ph.D. Wayne State 1985. An international comparison of the urban and regional quality of life: The U.S.A. and Europe.

KURT J. BERON, Ph.D. North Carolina-Chapel Hill 1985. Child support enforcement: A theoretical and empirical analysis.

LEONARD E. BURMAN, Ph.D. Minnesota 1985. A dynamic model of housing demand with transaction costs.

JAMES M. CATALDO, Ph.D. Columbia 1985. Employment and crime: A life cycle approach.

ISMAIL COLE, Ph.D. Pittsburgh 1986. The structure of aggregate production and the sources of growth in urban manufacturing: An empirical study of selected U.S. metropolitan areas.

HARRIET L. DULEEP, Ph.D. MIT 1986. Poverty and inequality in mortality.

ANASTASSIOS FRANGOS, Ph.D. Wayne State 1985. Classical location theory, spatial competition, and general equilibrium analysis.

PAUL J. GERTLER, Ph.D. Wisconsin-Madison 1985. Subsidies, quality, and regulation in the nursing home industry.

EDWARD GULLASON, Ph.D. Pennsylvania 1986. Essays on the private returns to schooling.

CHRISTINE J. HAGER, Ph.D. North Carolina State 1985. Demand for nutrient and nonnutrient components in household purchases of red meat, poultry, and fish products using a hedonic approach.

WILLIAM E. HERRIN, Ph.D. SUNY-Binghamton 1986. Ascertaining the validity of the standard urban model: An implicit markets approach.

CHRISTOPHER HOGAN, Ph.D. Northwestern 1986. Adverse selection in the health care market.

KATHRYN L. HOLLIDAY, Ph.D. Washington 1986. Informational efficiency of residential real estate markets: Study of Seattle housing market, 1973–80.

SARAH B. JENKINS, Ph.D. South Carolina 1985. Analysis of the age composition of the U.S. population in the demand for financial assets: An empirical study.

DONALD L. JOHNSON, Ph.D. Tennessee-Knoxville 1986. An analysis of the determinants of worker migration: Implications for the location of high technology industries.

SUKNAM KO, Ph.D. Illinois-Urbana 1985. Regional general equilibrium model.

FRANCIS T. LUI, Ph.D. Minnesota 1985. Three essays on the economics of corruption.

LINDA C. MCCARTHY, Ph.D. Purdue 1985. The effect of plea bargaining of the form of concessions: A theoretical and empirical analysis.

MICHAEL J. MCKEE, Ph.D. Carleton 1985. Aspects of competitive bureaus: Lessons from public sector schooling in Ontario and British Columbia.

ALAN D. MATHIOS, Ph.D. Pennsylvania 1985. Education and the variation in earnings.

STEVEN A. MEYER, Ph.D. Virginia 1985. Effects of an income maintenance program on job choice.

JOHN MULLAHY, Ph.D. Virginia 1985. Cigarette smoking: Habits, health concerns, and heterogeneous unobservables in a microeconometric analysis of consumer demand.

RICHARD P. NUMRICH, Ph.D. SUNY-Albany 1986. Urban spatial structure: A generalized varying coefficient model.

LYNN PAQUETTE, Ph.D. MIT 1985. The consumption and living arrangements of the elderly.

WANKYU PARK, Ph.D. Wisconsin-Madison 1985. An empirical analysis of the efficient provision of a public good.

PEDRO T. PEREIRA, Ph.D. Minnesota 1986. Marriage and divorce.

ELLEN P. ROCHE, Ph.D. Maryland 1986. The effect of property taxes on renters: Theoretical and empirical research on a model of dynamic incidence.

GLENN R. ROGERS, Ph.D. Wisconsin-Madison 1986. The theory of output-income multipliers with consumption linkages: An application to Mauritania.

MARIA J. SAN SEGUNDO GOMEZ DE CADINANOS, Ph.D. Princeton 1985. Emprical studies of quality.

TAPENDRA N. SINHA, Ph.D. Minnesota 1986. Unintended bequests: Theory and evidence.

BARBARA J. P. STREET, Ph.D. Virginia 1986. Implications of liability rules for physicians and hospitals.

SEOUNG HWAN SUH, Ph.D. Princeton 1985. Three essays on the theory of multiple cities.

GARY N. TOMPKINS, Ph.D. Western Ontario 1985. Competitive land markets in a nonmonocentric urban area.

PETER G. C. TOWNLEY, Ph.D. Queen's 1986. Provision of retirement incomes under political and informational restraints.

WILLIAM N. TRUMBULL, Ph.D. North Carolina-Chapel Hill 1985. Incorporating uncertainty in policy analysis with an application in criminal justice.

SOMCHAI VIRUNHAPHOL, Ph.D. SUNY-Binghamton 1986. The economics of urban revitalization.

NANCY L. WOLFF, Ph.D. Iowa State 1984. The distribution impact of the Social Security Program, 1962–72.

ANDREA L. ZIEGERT, Ph.D. North Carolina-Chapel Hill 1985. The effect of anticipated housing asset inflation on homeowner durable and nondurable good demand.

Not Classified

ALBERTO ALESINA, Ph.D. Harvard 1986. Macroeconomic policy in a two-party system.

EDWARD J. AMADEO, Ph.D. Harvard 1986. Keynes' principle of effective demand and its relationship to alternative theories of distribution and accumulation.

KOFI A. AMOATENG, Ph.D. Texas-Austin 1986. Essays on international debt: Theory and evidence.

NAUZER BALSARA, Ph.D. Columbia 1986. The role of currency baskets in international portfolio management.

LINDA BELL, Ph.D. Harvard 1986. Essays on labor market efficiency and comparative macroeconomic performance.

CIHAN BILGINSOY, Ph.D. Massachusetts-Amherst 1986. Monetarist and structuralist models of Turkish inflation, 1950–80.

STEVEN M. BLOOM, Ph.D. Harvard 1986. Employee ownership and firm performance.

DOWLAT R. BUDHRAM, Ph.D. Ohio State 1986. Sugarcane-based ethanol: Production possibilities and trade implications for Caribbean countries.

PATRICIA BYRNES, Ph.D. Southern Illinois-Carbondale 1985. Ownership and efficiency in the water supply industry: An application of the nonparametric programming approach to efficiency measurement.

JORGE CASTELLANOS, Ph.D. Columbia 1986. Essays on lending to developing countries.

CLARKE L. CAYWOOD, Ph.D. Wisconsin-Madison 1986. Toward a novel theory of political advertising regulation.

KIMBERLY CHRISTENSEN, Ph.D. Massachusetts-Amherst 1986. Accounting for the relative income gains of black females, 1955–82.

STEVEN M. COHN, Ph.D. Massachusetts-Amherst 1986. Political economy of nuclear power: 1946–82.

CIYATA COLEMAN, Ph.D. Southern Illinois-Carbondale 1985. Exports and spatial economic growth in Liberia: Effects of rubber and iron ore.

JOHN V. COLIAS, Ph.D. Texas-Austin 1985. A sectoral applied general equilibrium model of the U.S. economy.

EDDIE DEKEL, Ph.D. Harvard 1986. Essays on choices and beliefs in uncertainty and game theory.

MATHIAS DEWATRIPONT, Ph.D. Harvard 1986. On the theory of commitment with applications to the labor market.

STEVEN DICKEY, Ph.D. Southern Illinois-Carbondale 1986. Alternative models of the firms' demand for money.

MOHAMMED T. A. EL-RAJABI, Ph.D. Wisconsin-Madison 1986. The information content of announcing the capital investment decisions and the actual investment numbers.

MEHRDAD E. FARIMANI, Ph.D. Wisconsin-Madison 1986. Neoclassical investment theory: A reexamination.

SALAM K. FAYYAD, Ph.D. Texas-Austin 1986. Monetary asset component grouping and aggregation: An inquiry into the definition of money.

MERYL FINKEL, Ph.D. Harvard 1986. Overseas research and development by U.S. multinationals: Ownership structure decisions.

GUSTAVO H. B. FRANCO, Ph.D. Harvard 1986. Aspects of the economics of hyperinflation; and theoretical issues and historical studies of four European hyperinflations of the 1920's.

GERALD C. FRIEDMAN, Ph.D. Harvard 1986. Politics and unions: Government ideology and the labor movement in the United States and France, 1880–1914.

CAROL B. GAUMNITZ, Ph.D. Wisconsin-Madison 1986. An analysis of the Internal Revenue Services decision process with respect to Sections 511–513.

STUART N. GLOSSER, Ph.D. Texas-Austin 1985. The influence of credit market rate on the demand for housing and automobiles.

BERIK GUNSELI, Ph.D. Massachusetts-Amherst 1986. Women's employment in hand-woven carpet production in rural Turkey.

EROL HAKANOGLU, Ph.D. Harvard 1984. Studies in the isomorphism between resource allocation algorithms and electrical networks.

KIFFLE HENOCK, Ph.D. Massachusetts-Amherst 1986. The determinants of the economic policies of states in the Third World: Agrarian policies of the Ethiopian state, 1941–74.

PETER HIGGS, Ph.D. Harvard 1986. Applied general equilibrium economics: Data and applications.

JAMES R. HINES, Ph.D. Harvard 1986. Public goods and tax structure.

TEH-MING HUO, Ph.D. Rochester 1985. Essays on intertemporal monetary models.

DANIEL E. INGBERMAN, Ph.D. Carnegie-Mellon 1986. Spatial competition with imperfectly informed voters.

AVRAHAM KAMARA, Ph.D. Columbia 1986. Equilibrium and efficiency in futures markets with delivery risk.

AVERY KATZ, Ph.D. Harvard 1986. Essays in the economics of litigation.

PATRICK KEHOE, Ph.D. Harvard 1986. Macroeconomics in a world economy.

IN JOON KIM, Ph.D. Columbia 1986. Essays on the valuations of contingent claims.

THOMAS A. KOLENKO, Ph.D. Wisconsin-Madison 1986. Effects of individual-task congruence on performance, satisfaction, and tenure.

POK-SANG LAM, Ph.D. Harvard 1986. Two essays on consumer durables expenditures.

JOAN LEEDS, Ph.D. Columbia 1985. Demand for heterogeneous labor in the services industries; substitution wage output and technology elasticities.

SIEW MENG LEONG, Ph.D. Wisconsin-Madison 1986. Knowledge, adaptive selling, and salesperson effectiveness: An empirical investigation from a script-theoretic perspective.

THOMAS J. LINSMEIER, Ph.D. Wisconsin-Madison 1986. A debt covenant rationale for a market reaction to a mandated accounting change: The case of the investment tax credit.

STEPHEN A. LORENZEN, Ph.D. Texas-Austin 1986. Employment, earnings, and consumption strategies in urban Mexico.

ELAINE E. MCCRATE, Ph.D. Massachusetts-Amherst 1985. The growth of nonmarriage among U.S. women, 1954–83.

JAYASHREE MAHAJAN, Ph.D. Wisconsin-Madison 1986. Sales organization and effectiveness: A structure-contingency approach.

WEI-LIN MAO, Ph.D. Texas-Austin 1986. Trade, structure, and strategic planning: A study of the Taiwan economy using optimal control.

RANDALL MORCK, Ph.D. Harvard 1986. Essays on capital structure.

DONALD V. MOSER, Ph.D. Wisconsin-Madison 1986. The effects of interference, availability, and accounting information on investors' predictive judgments.

ABRAHAM MULUGETTA, Ph.D. Wisconsin-Madison 1986. Variability of returns of common stock of MNEs as a result of geographical segment information disclosure.

PAUL S. NELSON, Ph.D. Texas A&M 1985. Capital investment in natural monopolies under uncertainty with allocation rules.

SUNGBAILK OH, Ph.D. Wisconsin-Madison 1986. A multiperiod investigation of agency costs of debt and the impact of bond indenture provisions.

WILLIAM C. OLSON, Ph.D. Massachusetts-Amherst 1985. Concepts of class in economic theory: A critique and reformulation.

CAROL D. PETERSEN, Ph.D. Texas-Austin 1986. Intrastate inequalities in the financing and delivery of noneducational social services to children and youth: An assessment of the state of Nebraska.

GORDON S. POTTER, Ph.D. Wisconsin-Madison 1986. Accounting earnings announcements, fund concentration, and common stock returns.

JEAN L. PYLE, Ph.D. Massachusetts-Amherst 1985. Sex discrimination and public policy in an open economy.

DANNY QUAH, Ph.D. Harvard 1986. Essays in dynamic macroeconometrics.

STEVEN RAYMAR, Ph.D. Columbia 1985. Capital structure when earnings are mean-reverting.

JAMES B. REBITZER, Ph.D. Massachusetts-Amherst 1985. Long-term employment relations in the U.S. economy.

JOSEPH M. RICCIARDI, Ph.D. Texas-Austin 1985. Essays on the role of money and finance in economic development.

SUSAN ROSENBERG, Ph.D. Columbia 1986. The financial economics of mortgages.

TIMOTHY RUPINSKY, Ph.D. Harvard 1986. The economics of managerial compensation.

NARENDRA K. RUSTAGI, Ph.D. Ohio State 1986. A simulation analysis of the Indian crop insurance program.

ZULKIFLI ZENTERI, Ph.D. Ohio State 1985. An econometric analysis of the U.S. palm oil market.

DAVID SIMON, Ph.D. Columbia 1986. Monetary policy, currency substitution and the transmission of stocks from the United States to Canada, 1976–84: A rational expectations approach.

KENNETH WEILLER, Ph.D. Harvard 1986. Essays on the effects of fiscal policies in open economies.

STEPHEN N. WHITE, JR., Ph.D. Texas A&M 1985. The generalized Fechner-Thurstone direct utility function as an alternative basis for computation of true cost-of-living indexes.

BARBARA WIGET, Ph.D. Harvard 1986. The growth of female employment: Two economic hypotheses.

MICHAEL WILLOUGHBY, Ph.D. Columbia 1985. The determinants of technological advance in banking: A logit model of automatic teller machine adoption.

LORETTA S. WILSON, Ph.D. Massachusetts-Amherst 1985. A reappraisal of international trade theories: New dimensions in criticism and comparison.

MING-MING WU, Ph.D. Ohio State 1986. The demand analysis for four major milk and dairy products in the United States: An application of logistics.

MICHAEL ZALKIN, Ph.D. Massachusetts-Amherst 1986. Peasant's response to state grain policy in postrevolutionary Nicaragua, 1979–84.

NOTICE TO ALL GRADUATE DEPARTMENTS

PLEASE NOTE: The *American Economic Review* will no longer carry the annual list of doctoral dissertation recipients published by the AEA.

The *Journal of Economic Literature* will print the eighty-fourth list of recipients and titles of degrees conferred in economics in U.S. and Canadian universities and colleges for the academic year ending June 1987. This announcement is an invitation to send them information for the preparation of the list.

By June 30, 1987, please send individual 3×5 cards, conforming to the style shown below. Please use the *JEL* classification system that is printed in each issue of the *JEL*.

ALL ITEMS AND INFORMATION FOR THE 1987 LIST SHOULD BE SENT TO:

The Journal of Economic Literature, P.O. Box 7320, Oakland Station, Pittsburgh, PA 15213

JEL Classification No. ________

Name: LAST NAME IN CAPS: First Name, Initial

Institution Granting Degree: ____________________

Degree Conferred (Ph.D. or D.B.A.) __________ Year ____________

Dissertation Title: ____________________

THE AMERICAN ECONOMIC REVIEW

VOLUME LXXVI

BOARD OF EDITORS

Clive Bull
Michael R. Darby
Jacob A. Frenkel
Claudia D. Goldin
Philip E. Graves
George E. Johnson
John F. Kennan
Mervyn A. King
Meir Kohn
Paul Krugman
Bennett T. McCallum
Edgar O. Olsen
Alvin E. Roth
Steven Shavell
John B. Shoven
Susan Woodward
Leslie Young

MANAGING EDITOR

ORLEY ASHENFELTER

CO-EDITORS

ROBERT H. HAVEMAN

JOHN G. RILEY

JOHN B. TAYLOR

THE AMERICAN ECONOMIC ASSOCIATION

Executive Office: Nashville, Tennessee

Editorial Office: 209 Nassau Street, Princeton, NJ 08542-4607

Copyright 1986

All Rights Reserved

AMERICAN ECONOMIC ASSOCIATION

CONTENTS OF ARTICLES AND SHORTER PAPERS

C. P. Kindleberger: International Public Goods without International Government 1

E. P. Lazear: Retail Pricing and Clearance Sales 14

M. Dotsey and R. G. King: Informational Implications of Interest Rate Rules 33

S. A. Lippman and J. J. McCall: An Operational Measure of Liquidity 43

W. J. Ethier: Illegal Immigration: The Host-Country Problem 56

M. Obstfeld: Rational and Self-Fulfilling Balance-of-Payments Crises 72

B. Bernanke: Employment, Hours, and Earnings in the Depression: An Analysis of Eight Manufacturing Industries 82

M. C. Lovell: Tests of the Rational Expectations Hypothesis 110

J. A. Miron: Financial Panics, the Seasonality of the Nominal Interest Rate, and the Founding of the Fed 125

Z. Griliches: Productivity, *R&D* and Basic Research at the Firm Level in the 1970's 141

C. R. Knoeber: Golden Parachutes, Shark Repellents, and Hostile Tender Offers 155

D. G. Bivin: Inventories and Interest Rates: A Critique of the Buffer Stock Model 168

D. Dollar: Technological Innovation, Capital Mobility, and the Product Cycle in North-South Trade 177

R. Ram: Government Size and Economic Growth: A New Framework and Some Evidence from Cross-Section and Time-Series Data 191

R. Cantor: A Macroeconomic Model with Auction Markets and Nominal Contracts 204

F. R. Kaen and R. E. Rosenman: Predictable Behavior in Financial Markets: Some Evidence in Support of Heiner's Hypothesis 212

Y. Benjamini and Y. Benjamini: The Choice Among Medical Insurance Plans 221

S. Schwab: Is Statistical Discrimination Efficient? 228

A. S. Holland: Wage Indexation and the Effect of Inflation Uncertainty on Employment: An Empirical Analysis 235

R. N. Anthony: Accounting Rates of Return: Note 244

M. Gallagher: The Inverted Fisher Hypothesis: Additional Evidence 247

J. G. Cullis and P.R. Jones: Rationing by Waiting Lists: An Implication 250

S. E. Plaut: Implicit Contracts in the Absence of Enforcement: Note 257

L. Arvan and J. K. Brueckner: Efficient Contracts in Credit Markets Subject to Interest Rate Risk: An Application of Raviv's Insurance Model 259

D. W. Blair, R. L. Cottle, and M. S. Wallace: Faculty Ratings of Major Economics Departments by Citations: An Extension 264

A. R. Schwartz, M. S. Cohen, and D. R. Grimes: Structural/Frictional vs. Deficient Demand Unemployment: Comment 268

K. G. Abraham: Reply 273

C. E. Bohanon and T. N. Van Cott: Labor Supply and Tax Rates: Comment 277

F. Ghavari: Comment 280

J. Gwartney and R. L. Stroup: Reply 284

G. Briden and J. Zedella: Social Security and Household Savings: Comment 286

E. Koskela and M. Virén: Reply 289

F. Modigliani: Life Cycle, Individual Thrift, and the Wealth of Nations 297

C. D. Romer: Is the Stabilization of the Postwar Economy a Figment of the Data? 314

T. M. Stoker: The Distributional Welfare Effects of Rising Prices in the United States: The 1970's Experience 335

E. Wicker: Terminating Hyperinflation in the Dismembered Habsburgh Monarchy 350

A. J. Oswald: Unemployment Insurance and Labor Contracts under Asymmetric Information: Theory and Facts 365

R. Rob: The Design of Procurement Contracts 378

J. J. Rotemberg and G. Saloner: A Supergame-Theoretic Model of Price Wars during Booms 390

D. E. Bloom and C. L. Cavanagh: An Analysis of the Selection of Arbitrators 408

J. S. Tracy: An Investigation into the Determinants of U.S. Strike Activity 423

H. E. Peters: Marriage and Divorce: Informational Constraints and Private Contracting 437

W. R. Johnson and J. Skinner: Labor Supply and Marital Separation 455

M. R. Rozenzweig and K. I. Wolpin: Evaluating the Effects of Optimally Distributed Public Programs: Child Health and Family Planning Interventions 470

T. A. Marsh and R. C. Merton: Dividend Variability and Variance Bounds Tests for the Rationality of Stock Market Prices 483

R. J. Shiller: The Marsh-Merton Model of Managers' Smoothing of Dividends 499

T. F. Cooley and S. F. LeRoy: What Will Take the Con Out of Econometrics? A Reply to McAleer, Pagan, and Volker 504

C. A. Holt: Preference Reversals and the Independence Axiom 508

J. H. Cassing and A. L. Hillman: Shifting Comparative Advantage and Senescent Industry Collapse . . . 516

E. Applebaum and E. Katz: Measures of Risk Aversion and Comparative Statics of Industry Equilibrium . . . 524

R. Wintrobe and A. Breton: Organizational Structure and Productivity . . . 530

U. Kohli: Robert Giffen and the Irish Potato: Note 539

M. D. Ramirez: Marx and Malthusianism: Comment . . . 543

S. Hollander: Reply . . . 548

D. M. Betson and D. Greenberg: Labor Supply and Tax Rates: Comment . . . 551

J. D. Gwartney and R. L. Stroup: Reply . . . 557

M. Cherkes, J. Friedman, and A. Spivak: The Disinterest in Deregulation: Comment . . . 559

R. E. McCormick, W. F. Shughart II, and R. D. Tollison: Reply . . . 564

M. P. Todaro: Internal Migration and Urban Employment: Comment . . . 566

W. E. Cole and R. D. Sanders: Reply . . . 570

W. G. Dewald, J. G. Thursby, and R. G. Anderson: Replication in Empirical Economics: *The Journal of Money, Credit and Banking Project* . . . 587

R. J. LaLonde: Evaluating the Econometric Evaluations of Training Programs with Experimental Data . . . 604

G. W. Evans: A Test for Speculative Bubbles and the Sterling–Dollar Exchange Rate: 1981–84 621

D. W. Carlton: The Rigidity of Prices . . . 637

J. M. Poterba, J. J. Rotemberg, and L. H. Summers: A Tax-Based Test for Nominal Rigidities . . . 659

R. B. Barsky, N. G. Mankiw, and S. P. Zeldes: Ricardian Consumers with Keynesian Propensities . . . 676

C. E. Walsh: In Defense of Base Drift . . . 692

S. Rosen: Prizes and Incentives in Elimination Tournaments . . . 701

R. K. Sah and J. E. Stiglitz: The Architecture of Economic Systems: Hierarchies and Polyarchies . . . 716

D. Kahneman, J. L. Knetsch, and R. Thaler: Fairness as a Constraint on Profit Seeking: Entitlements in the Market . . . 728

T. Bresnahan: Measuring the Spillovers from Technical Advance: Mainframe Computers in Financial Services . . . 742

M. Gisser: Price Leadership and Welfare Losses in U.S. Manufacturing . . . 756

P. K. Robins: Child Support, Welfare Dependency, and Poverty . . . 768

B. D. Bernheim: On the Voluntary and Involuntary Provision of Public Goods . . . 789

R. A. Margo: Educational Achievement in Segregated School Systems: The Effects of "Separate-but-Equal" . . . 794

R. A. Driskill and S. M. Sheffrin: Is Price Flexibility Destabilizing? . . . 802

J. D. Hamilton and M. A. Flavin: On the Limitations of Government Borrowing: A Framework for Empirical Testing . . . 808

E. W. Bond and L. Samuelson: Tax Holidays as Signals . . . 820

G. Loury and T. R. Lewis: On the Profitability of Interruptible Supply . . . 827

M. Crain, D. Leavens, and R. D. Tollison: Final Voting in a Legislature . . . 833

B. Wernerfelt: Product Line Rivalry: Note . . . 842

B. A. Polkinghorn: An Unpublished Letter from Malthus to Jane Marcet, January 22, 1833 . . . 845

R. L. Greenfield and L. B. Yeager: Competitive Payments Systems: Comment . . . 848

L. H. White: Reply . . . 850

T. Sueyoshi and P. C. Anselmo: The Evans and Heckman Subadditivity Test: Comment . . . 854

J. Yinger: Measuring Racial Discrimination with Fair Housing Audits: Caught in the Act . . . 881

J. H. Kagel and D. Levin: The Winner's Curse and Public Information in Common Value Auctions 894

P. Rey and J. Tirole: The Logic of Vertical Restraints . . . 921

J. Farrell and G. Saloner: Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation . . . 940

J. A. Brander and T. R. Lewis: Oligopoly and Financial Structure: The Limited Liability Effect . . . 956

M. Aoki: Horizontal vs. Vertical Information Structure of the Firm . . . 971

A. B. Jaffe: Technological Opportunity and Spillovers of *R&D*: Evidence from Firms' Patents, Profits, and Market Value . . . 984

J. R. Markusen: Explaining the Volume of Trade: An Eclectic Approach . . . 1002

P. M. Garber: Nominal Contracts in a Bimetallic Standard . . . 1012

J. B. De Long and L. H. Summers: Is Increased Price Flexibility Stabilizing? . . . 1031

D. Card: Efficient Contracts with Costly Adjustment: Short-Run Employment Determination for Airline Mechanics . . . 1045

W. J. Baumol: Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show 1072

R. H. Gordon: Taxation of Investment and Savings in a World Economy: The Certainty Case 1086

D. Lam: The Dynamics of Population Growth, Differential Fertility, and Inequality . . . 1103

G. Johnson and G. Solon: Estimates of the Direct Effects of Comparable Worth Policy 1117

R. J. Ruffin and F. Rassekh: The Role of Foreign Direct Investment in U.S. Capital Outcomes 1126

J. R. Melvin: The Nonequivalence of Tariffs and Import Quotas 1131

E. Lichtenberg and D. Zilberman: The Welfare Economics of Price Supports in the U.S. Agriculture 1135

J. M. Poterba and L. H. Summers: The Persistence of Volatility and Stock Market Fluctuations 1142

C. L. Schultze: The Cyclical Flexibility of Wages 1152

R. D. Milbourne and D. F. Richards: A New View of the Federal Debt and Budget Deficits: Comment 1154

R. Eisner and P. J. Pieper: Reply 1156

J. R. Barth, G. Iden, and F. S. Russek: Government Debt, Government Spending, and Private Sector Behavior: Comment 1158

F. Modigliani and A. Sterling: Comment 1168

R. C. Kormendi and P. Meguire: Reply 1180

A. G. Blomqvist: The Economics of Price Scissors: Comment 1188

M. R. Carter: Comment 1192

R. K. Sah and J. E. Stiglitz: Reply 1195

E. Alvi: The Production Process in a Competitive Economy: Comment 1200

S. Bowles: Reply 1203

W. G. Shepherd: Tobin's *q* and the Structure Performance Relationship: Comment 1205

M. Smirlock, T. Gilligan, and W. Marshall: Reply 1211

E. K. Browning: Pechman's Tax Incidence Study: A Note on the Data 1214

J. Pechman: A Response 1219

R. A. Margo: Race and Human Capital: Comment 1221

J. P. Smith: Reply 1225

◆

CONTENTS OF THE PAPERS AND PROCEEDINGS

Richard T. Ely Lecture

H. Stein: The Washington Economics Industry 1

Economic Issues in the Arts

W. J. Baumol: Unnatural Value: or Art Investment as Floating Crap Game 10

D. Netzer: Dance In New York: Market and Subsidy Changes 15

J. H. Gapinski: The Lively Arts as Substitutes for the Lively Arts 20

Supply-Side Economics: What Remains?

M. Feldstein: Supply-Side Economics: Old Truths and New Claims 26

L. Chimerene and R. M. Young: Economic Surprises and Messages of the 1980's 31

G. M. von Furstenberg and R. J. Green: Supply-Side Modeling from Bits and Pieces 37

Occupations and Labor Markets: A Critical Evaluation

W. T. Bielby and J. N. Baron: Sex Segregation Within Operations 43

P. B. Doeringer: Internal Labor Markets and Noncompeting Groups 48

M. A. Ferber, C. A. Green, and J. L. Spaeth: Work Power and Earnings of Women and Men . . . 53

Politics and Economic Policies

J. E. Alt: Party Strategies, World Demand, and Unemployment: The Political Economy of Economic Activity in Western Industrial Nations 57

K. A. Chrystal and D. A. Peel: What Can Economics Learn from Political Science, and Vice Versa? 62

D. A. Hibbs, Jr.: Political Parties and Macroeconomic Policies and Outcomes in the United States 66

H. W. Chappell, Jr. and W. R. Keech: Party Differences in Macroeconomic Policies and Outcomes 71

Developing Country Policy Responses To Exogenous Shocks

B. Belassa: Policy Responses to Exogenous Shocks in Developing Countries 75

M. J. Fry and D. M. Lilien: Monetary Policy Responses to Exogenous Shocks 79

M. S. Khan: Developing Country Exchange Rate Policy Responses to Exogenous Shocks 84

V. Tanzi: Fiscal Policy Responses to Exogenous Shocks in Developing Countries 88

Unions in Decline: Causes and Consequences

R. B. Freeman: The Effect of the Union Wage Differential on Management Opposition and Union Organizing Success 92

R. Edwards and P. Swaim: Union-Nonunion Earnings Differentials and the Decline of Private-Sector Unionism 97

P. Linneman and M. L. Wachter: Rising Union Premiums and the Declining Boundaries Among Noncompeting Groups 103

Economic Policy and the Theory of the Firm: New Perspectives

T. G. Krattenmaker and S. C. Salop: Competition and Cooperation in the Market for Exclusionary Rights 109

O. E. Williamson: Transforming Merger Policy: The Pound of New Perspectives 114

Budget Reform and the Theory of Fiscal Federalism

M. Olson: Toward a More General Theory of Government Structure 120

D. R. Feenberg and H. S. Rosen: The Interaction of State and Federal Tax Systems: The Impact of State and Local Tax Deductibility . . . 126

J. M. Quigley and D. L. Rubinfeld: Budget Reform and the Theory of Fiscal Federalism . . 132

Roundtable on Economic Education: Increasing the Public's Understanding of Economics

A. Rees: The Marketplace of Economic Ideas . . . 138

L. Silk: Communicating Economic Ideas and Controversies 141

M. A. MacDowell: Increasing the Public's Understanding of Economics: What Can We Expect from the Schools? 145

W. L. Hansen: What Knowledge is Most Worth Knowing—For Economics Majors? 149

Economic Issues in U.S. Infrastructure Investment

T. E. Keeler: Public Policy and Productivity in the Trucking Industry: Some Evidence of Highway Investments, Deregulation, and the 55 MPH Speed Limit 153

G. E. Peterson: Urban Road Reinvestment: The Effects of External Aid 159

K. A. Small and C. Winston: Efficient Pricing and Investment Solutions to Highway Infrastructure Needs 165

The Soviet Growth Slowdown: Three Views

S. Gomulka: Soviet Growth Slowdown: Duality, Maturity, and Innovation 170

P. Desai: Soviet Growth Retardation 175

V. Kontorovich: Soviet Growth Slowdown: Econometric vs. Direct Evidence 181

R&D, Innovation, the Public Policy

R. R. Nelson: Institutions Supporting Technical Advance in Industry 186

E. Mansfield: The *R&D* Tax Credit and Other Technology Policy Issues 190

H. Grabowski and J. Vernon: Longer Patents for Lower Imitation Barriers: The 1984 Drug Act 195

R. C. Levin: A New Look at the Patent System 199

The Monetary-Fiscal Policy Mix: Implications for Macroeconomic Performance

A. F. Brimmer and A. Sinai: The Monetary-Fiscal Policy Mix: Implications for the Short Run 203
F. C. Ribe and W. J. Beeman: The Monetary-Fiscal Mix and Long-Run Growth in an Open Economy 209
J. Tobin: The Monetary-Fiscal Mix: Long-Run Implications 213

Welfare Reform: New Research and Policy Developments

R. Moffit: Work Incentives in the AFDC System: An Analysis of the 1981 Reforms 219
D. Friedlander, B. Goldman, J. Gueron, and D. Long: Initial Findings from the Demonstration of State Work/Welfare Initiatives . . 224
T. Fraker, B. Devaney, and E. Cavin: An Evaluation of the Effect of Cashing Out Food Stamps on Food Expenditures 230

Changes in Wage Norms

A. Lindbeck and D. J. Snower: Wage Setting, Unemployment, and Insider-Outsider Relations 235
M. L. Wachter: Union Wage Rigidity: The Default Settings of Labor Law 240
G. L. Perry: Shifting Wage Norms and their Implications 245
D. J. B. Mitchell: Union vs. Nonunion Wage Norm Shifts 249

Economic Issues in Immigration Policy

B. R. Chiswick: Illegal Aliens: A Preliminary Report on an Employee-Employer Survey 253
W. J. Ethier: Illegal Immigration 258
F. L. Rivera-Batiz: Can Border Industries Be a Substitute for Immigration? 263

The Political Economy of Outer Space

L. R. Cohen and R. G. Noll: Government *R&D* Programs for Commercializing Space 269
J. O. Ledyard: Incentive Compatible Space Station Pricing 274
M. K. Macauley: Out of Space? Regulation and Technical Change in Communications Satellites 280

Siting of Hazardous Facilities

R. C. Mitchell and R. T. Carson: Property Rights, Protest, and the Siting of Hazardous Waste Facilities 285
V. K. Smith and W. H. Desvousges: Asymmetries in the Valuation of Risk and the Siting of Hazardous Waste Disposal Facilities 291
H. Kunreuther and P. R. Kleindorfer: A Sealed-Bid Auction Mechanism for Siting Noxious Facilities 295

Regional Growth Patterns: Trends, Prospects and Policy Implications

B. Chinitz: The Regional Transformation of the American Economy 300
B. H. Stevens and G. I. Treyz: A Multiregional Model Forecast for the United States Through 1995 304
C. L. Leven: Analysis and Policy Implications of Regional Decline 308

The Market for Corporate Control

H. Demsetz: Corporate Control, Insider Trading, and Rates of Return 313
M. Hirschey: Mergers, Buyouts and Fakeouts . . . 317
M. C. Jensen: Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers 323

The International Dimensions of Fiscal Policies

J. A. Frenkel and A. Razin: The International Transmission and Effects of Fiscal Policies 330
J. Sachs: The Uneasy Case for Greater Exchange Rate Coordination 336
M. Feldstein: U.S. Budget Deficits and the European Economies: Resolving the Political Economy Puzzle 342

Do Government Programs Close the Racial Gap?

E. D. Carson: The Black Underclass Concept: Self-Help vs. Government Intervention 347
W. Vroman: Transfer Payments, Sample Selection and Male Black-White Earnings Differences 351
J. M. Culp, Jr.: Federal Courts and the Enforcement of Title VII 355
J. S. Leonard: What was Affirmative Action? . . . 359

Equity Between the Sexes in Economic Participation

E. Sorensen: Implementing Comparable Worth: A Survey of Recent Job Evaluation Studies 364
M. White: Sex Differences in Urban Commuting Patterns 368
N. L. Maxwell and R. J. D'Amico: Employment and Wage Effects of Involuntary Job Separation: Male-Female Differences 373
N. Zalokar: Generational Differences in Female Occupational Attainment—Have the 1970's Changed Women's Opportunities? 378

Oligopolistic Markets with Price-Setting Firms

E. Maskin: The Existence of Equilibrium with Price-Setting Firms 382
B. Allen and M. Hellwig: Price-Setting Firms and the Oligopolistic Foundations of Perfect Competition 387
J. Sutton: Vertical Product Differentiation: Some Basic Themes 393

Government Policy and Poverty

F. S. Levy and R. C. Michel: Work for Welfare: How Much Good Will It Do? 399
S. Danziger and P. Gottschalk: Do Rising Tides Lift All Boats? The Impact of Secular and Cyclical Changes on Poverty 405

Distinguished Lecture on Economics in Government

G. Jaszi: An Economic Accountant's Audit 411

CONTRIBUTORS TO ARTICLES AND SHORTER PAPERS

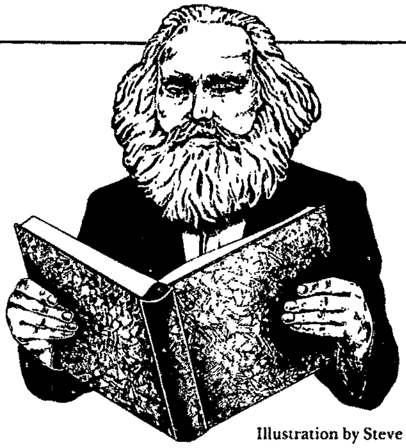
Abraham, K. G. 273
Alvi, E. 1200
Anderson, R. G. 587
Anselmo, P. C. 854
Anthony, R. N. 244
Aoki, M. 971
Applebaum, E. 524
Arvan, L. 259
Barsky, R. B. 676
Barth, J. R. 1158
Baumol, W. J. 1072
Benjamini, Y. 221
Benjamini, Y. 221
Bernanke, B. 82
Bernheim, B. D. 789
Betson, D. M. 551
Bivin, D. G. 168
Blair, D. W. 259
Blomqvist, A. G. 1188
Bloom, D. E. 408
Bohanon, C. E. 277
Bond, E. W. 820
Bowles, S. 1203
Brander, J. A. 956
Bresnahan, T. 742
Breton, A. 530
Briden, G. 286
Browning, E. K. 1214
Cantor, R. 204
Card, D. 1045
Carlton, D. W. 637
Carter, M. R. 1192
Cassing, J. H. 516
Cavanagh, C. L. 408
Cherkes, M. 559
Cohen, M. S. 268
Cole, W. E. 570
Cooley, T. F. 504
Cottle, R. L. 264
Crain, M. 833
Cullis, J. G. 250
De Long, J. B. 1031
Dewald, W. G. 587
Dollar, D. 177
Dotsey, M. 33
Driskill, R. A. 802
Eisner, R. 1156
Ethier, W. J. 56
Evans, G. W. 621
Farrell, J. 940
Flavin, M. A. 808
Friedman, J. 559
Gahvari, F. 280
Gallagher, M. 247
Garber, P. M. 1012
Gilligan, T. 1211
Gisser, M. 756
Gordon, R. H. 1086
Greenberg, D. 551
Greenfield, R. L. 848
Griliches, Z. 141
Grimes, D. R. 268
Gwartney, J. 284, 557
Hamilton, J. D. 808
Hillman, A. L. 516
Holland, A. S. 235
Hollander, S. 548
Holt, C. A. 508
Iden, G. 1158
Jaffe, A. B. 984
Johnson, G. 1117
Johnson, W. R. 455
Jones, P. R. 250
Kaen, F. R. 212
Kagel, J. H. 894
Kahneman, D. 728
Katz, E. 524
Kindleberger, C. P. 1
King, R. 33
Knetsch, J. L. 728
Knoeber, C. R. 155
Kohli, U. 539
Kormendi, R. C. 1180
Koskela, E. 289
LaLonde, R. J. 604
Lam, D. 1103
Lazear, E. P. 14
Leavens, D. 833
LeRoy, S. 504
Levin, D. 894
Lewis, T. R. 827, 956
Lichtenberg, E. 1135
Lippman, S. A. 43
Loury, G. C. 827
Lovell, M. C. 110
McCall, J. J. 43
McCormick, R. E. 564
Mankiw, N. G. 676
Margo, R. A. 794, 1221
Markusen, J. R. 1002
Marsh, T. A. 483
Marshall, W. 1211
Meguire, P. 1180
Melvin, J. R. 1131
Merton, R. C. 483
Milbourne, R. D. 1154
Miron, J. A. 125
Modigliani, F. 297, 1168
Obstfeld, M. 72
Oswald, A. J. 365
Pechman, J. A. 1219
Peters, H. E. 437
Pieper, P. J. 1156
Plaut, S. E. 257
Polkinghorn, B. A. 845
Poterba, J. M. 659, 1142
Ram, R. 191
Ramirez, M. D. 543
Rassekh, F. 1126
Rey, P. 921

Richards, D. J. 1154
Rob, R. 378
Robins, P. K. 768
Romer, C. 314
Rosen, S. 701
Rosenman, R. E. 212
Rosenzweig, M. R. 470
Rotemberg, J. J. 390, 659
Ruffin, R. J. 1126
Russek, F. S. 1158
Sah, R. K. 716, 1195
Saloner, G. 390, 940
Samuelson, L. 820
Sanders, R. D. 570
Schultze, C. L. 1152
Schwab, S. 228
Schwartz, A. R. 268
Sheffrin, S. M. 802
Shepherd, W. G. 1205
Shiller, R. J. 499
Shughart, W. F. II 564
Skinner, J. 455
Smirlock, M. 1211
Smith, J. P. 1225
Solon, G. 1117
Spivak, A. 559
Sterling, A. 1168
Stiglitz, J. E. 716, 1195
Stoker, T. M. 335
Stroup, R. L. 284, 557
Sueyoshi, T. 854
Summers, L. H. 659, 1031, 1142
Thaler, R. 728
Thursby, J. G. 587
Tirole, J. 921
Todaro, M. P. 566
Tollison, R. D. 564, 883
Tracy, J. S. 423
Van Cott, T. N. 277
Virén, M. 289
Wallace, M. 264
Walsh, C. E. 692
Wernerfelt, B. 842
White, L. H. 850
Wicker, E. 350
Wintrobe, R. 530
Wolpin, K. I. 470
Yeager, L. B. 848
Yinger, J. 881
Zedella, J. 286
Zeldes, S. P. 676
Zilberman, D. 1135

CONTRIBUTORS TO PAPERS AND PROCEEDINGS

Allen, B. 387
Alt, J. E. 57
Balassa, B. 75
Baron, J. N. 43
Baumol, W. J. 10
Beeman, W. J. 209
Bielby, W. T. 43
Brimmer, A. F. 203
Carson, E. D. 347
Carson, R. T. 285
Cavin, E. 230
Chappell, H. W., Jr. 71
Chimerine, L. 31
Chinitz, B. 300
Chiswick, B. R. 253
Chrystal, K. A. 62
Cohen, L. R. 269
Culp, J. M., Jr. 355
Danziger, S. 405
D'Amico, R. J. 373
Demsetz, H. 313
Desai, P. 175
Desvousges, W. H. 291
Devaney, B. 230
Doeringer, P. B. 48
Edwards, R. 97
Ethier, W. J. 258
Feenberg, D. R. 126
Feldstein, M. 26, 342
Ferber, M. A. 53
Fraker, T. M. 230
Freeman, R. B. 92
Frenkel, J. A. 330
Friedlander, D. 224
Fry, M. J. 79
Gapinski, J. H. 20
Goldman, B. 224
Gomulka, S. 170
Gottschalk, P. 405
Grabowski, H. 195
Green, C. A. 53
Green, R. J. 37
Gueron, J. 224
Hansen, W. L. 149
Hellwig, M. 387
Hibbs, D. A., Jr. 66
Hirschey, M. 317
Jaszi, G. 411
Jensen, M. C. 323
Keech, W. R. 71
Keeler, T. E. 153
Khan, M. S. 84
Kontorovich, V. 181
Kleindorfer, P. R. 295
Krattenmaker, T. G. 109
Kunreuther, H. 295
Ledyard, J. O. 274
Leonard, J. S. 359
Leven, C. L. 308
Levin, R. C. 199

Levy, F. S. 399
Lilien, D. M. 79
Lindbeck, A. 235
Linneman, P. 103
Long, D. 224
Macauley, M. K. 280
MacDowell, M. A. 145
Mansfield, E. 190
Maskin, E. 382
Maxwell, N. L. 373
Michel, R. C. 399
Mitchell, D. J. B. 249
Mitchell, R. C. 285
Moffitt, R. 219
Nelson, R. R. 186
Netzer, D. 15
Noll, R. G. 269
Olson, M. 120
Peel, D. A. 62
Perry, G. L. 245
Peterson, G. E. 159
Quigley, J. M. 132
Razin, A. 330
Rees, A. 138
Ribe, F. C. 209
Rivera-Batiz, F. L. 263
Rosen, H. S. 126
Rubinfeld, D. L. 132
Sachs, J. 336
Salop, S. C. 109
Silk, L. 141
Sinai, A. 203
Small, K. A. 165
Smith, V. K. 291
Snower, D. J. 235
Sorensen, E. 364
Spaeth, J. L. 53
Stein, H. 1
Stevens, B. H. 304
Sutton, J. 393
Swaim, P. 97
Tanzi, V. 88
Tobin, J. 213
Treyz, G. I. 304
Vernon, J. 195
von Furstenberg, G. M. 37
Vroman, W. 351
Wachter, M. L. 103, 240
White, M. 368
Williamson, O. E. 114
Winston, C. 165
Young, R. M. 31
Zalokar, N. 378



New titles from Harvard University Press

Illustration by Steve Damaskos.

Understanding Capital
Marx's Economic Theory
Duncan K. Foley
In a brilliantly lucid introduction to Marxist economic theory, Duncan Foley builds an understanding of the theory systematically, from first principles through the definition of central concepts to the development of important applications. All of the topics in the three volumes of *Capital* are included, providing the reader with a complete view of Marxist economics.
$20.00

The Logic and Limits of Bankruptcy Law
Thomas H. Jackson
Legal scholar Thomas H. Jackson identifies the underlying principles of bankruptcy law and develops an economic analysis of its main problems—a framework that permits him to view the field as a whole rather than as a collection of disparate policies and historical artifacts.
"I cannot think of any book on bankruptcy that is comparable to this in breadth or theoretical sophistication."—Anthony Kronman, Yale Law School
$25.00

The Control Revolution
Technological and Economic Origins of the Information Society
James R. Beniger
"An exquisite job of synthesis. This is one of those seminal works that change the way not only scholars but people in general look at the work."—Anthony Oettinger, Harvard University
$25.00

Medical Care, Medical Costs
The Search for a Health Insurance Policy
Rashi Fein
In a clearly written book that is both history and analysis, Fein makes an elegant plea for equity in the shaping of medical care delivery.
"An exceptionally well-written book, more cogently argued than anything extant. Sophisticated and sensible . . . the non-specialist can follow and enjoy reading it."—Eli Ginsberg
$20.00

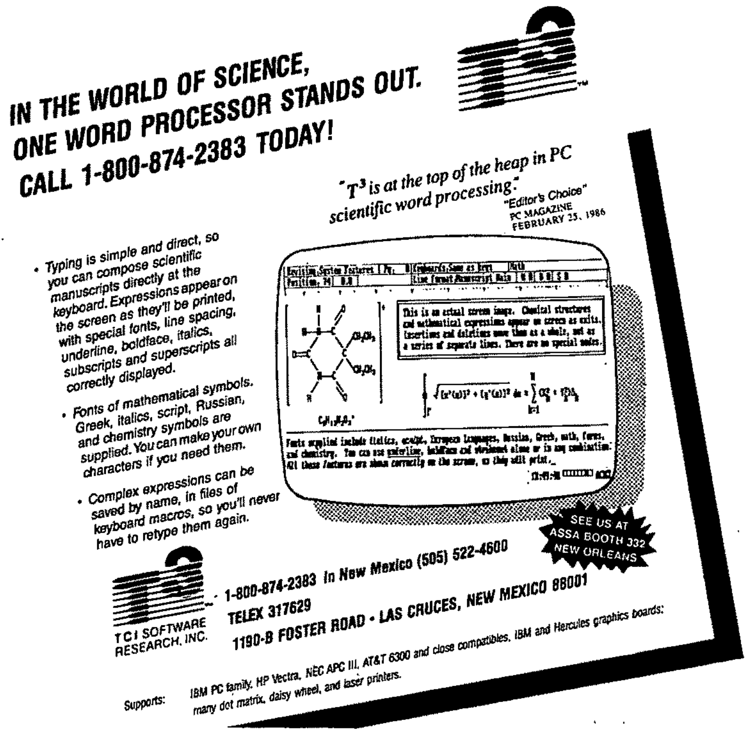
Paperbacks

Winner of the 1985 Pulitzer Prize in History

Prophets of Regulation
Thomas K. McCraw
"A novel, stimulating approach . . . McCraw's effort to clarify the nature of regulation by carefully integrating biography, history of ideas, and regulatory strategy pays handsome dividends."—*Washington Post Book World*
Belknap
$18.95 paper

Reality and Rhetoric
Studies in the Economics of Development
P.T. Bauer
"Peter Bauer is the world's foremost authority on economic development . . . In *Reality and Rhetoric* he destroys many popular myths about economic development . . . His scholarship is devastating, and it will leave sensitive readers with a sense of outrage at development economists whose ignorant meddling has devastated poor and struggling people."
—*Wall Street Journal*
$6.95 paper

Harvard
UNIVERSITY PRESS
79 GARDEN ST., CAMBRIDGE, MASSACHUSETTS 02138



IN THE WORLD OF SCIENCE,
ONE WORD PROCESSOR STANDS OUT.
CALL 1-800-874-2383 TODAY!

"T³ is at the top of the heap in PC scientific word processing."
"Editor's Choice"
PC MAGAZINE
FEBRUARY 25, 1986

- Typing is simple and direct, so you can compose scientific manuscripts directly at the keyboard. Expressions appear on the screen as they'll be printed, with special fonts, line spacing, underline, boldface, italics, subscripts and superscripts all correctly displayed.
- Fonts of mathematical symbols, Greek, italics, script, Russian, and chemistry symbols are supplied. You can make your own characters if you need them.
- Complex expressions can be saved by name, in files of keyboard macros, so you'll never have to retype them again.

SEE US AT ASSA BOOTH 332 NEW ORLEANS

TCI SOFTWARE RESEARCH, INC.

1-800-874-2383 In New Mexico (505) 522-4600
TELEX 317629
1190-B FOSTER ROAD • LAS CRUCES, NEW MEXICO 88001

Supports: IBM PC family, HP Vectra, NEC APC III, AT&T 6300 and close compatibles, IBM and Hercules graphics boards; many dot matrix, daisy wheel, and laser printers.

INTERNATIONAL MONETARY FUND
STAFF PAPERS

The latest issue of IMF Staff Papers contains studies by the Fund's staff on international monetary and financial problems. Among the articles included in the December 1986 issue are:

Output and Unanticipated Money with Imported Intermediate Goods and Foreign Exchange Rationing ....
by Ajai Chopra and Peter Montiel

Labor Market Disequilibrium and the Scope for Work-Sharing: A Case Study of the Netherlands ....
by Klaus-Walter Riechel

MINIMOD: Specifications and Simulation Results ....
by Richard D. Haas and Paul Masson

On Target Zones ....
by Jacob A. Frenkel and Morris Goldstein

Profits Theory and Profits Taxation ....
by Edmund S. Phelps

Subscription rates: US$15.00 per volume, US$4.00 for a single issue. Special rates for university libraries, faculty, and students: US$4.50 per volume.

Available from: Publications Unit, Box E-290 • International Monetary Fund
700 19th Street, N.W. • Washington, D.C. 20431, U.S.A.
Telephone (202) 623–7430

CHICAGO

ECONOMIC ADJUSTMENT AND EXCHANGE RATES IN DEVELOPING COUNTRIES

Edited by **Sebastian Edwards** *and* **Liaquat Ahamed**

Leading economists tackle various exchange-rate issues and their policy implications. Contributors: Sweder van Wijnbergen, Michael Mussa, Joshua Aizenman, Jacob A. Frenkel, Rudiger Dornbusch, Guillermo A. Calvo, Maurice Obstfeld, Sebastian Edwards, William H. Branson, Louka T. Katseli, Jorge Braga de Macedo, Arnold C. Hamberger, and commentators.

Cloth $48.00 456 pages NBER Conference Report

THE AMERICAN BUSINESS CYCLE

Continuity and Change

Edited by **Robert J. Gordon**

In the past decade the American economy went through a business cycle that included its worst peacetime inflation in history and its highest unemployment rate since the Great Depression. This volume reexamines the nature and causes of such cycles, especially in light of events such as the Korean and Vietnam wars and the energy crisis, and provides a broad historical view of changes in United States business-cycle behavior.

Cloth $78.00 896 pages
NBER Studies in Business Cycles, Volume 25

LONG-TERM FACTORS IN AMERICAN ECONOMIC GROWTH

Edited by **Stanley L. Engerman** *and* **Robert E. Gallman**

These studies of the historical process of economic change in nineteenth- and twentieth-century United States, Canada, and the British West Indies represent a major step toward understanding the sources of economic growth. The contributors discuss national product; capital stock and the distribution of wealth; fertility, health, and mortality; labor supply; and agriculture, manufacturing, transportation, and government.

Cloth $69.00 (est.) 696 pages (est.)
NBER Studies in Income and Wealth, Volume 51

Now in paper

MACROECONOMICS

A Neoclassical Introduction

Merton H. Miller *and* **Charles W. Upton**

With a new Preface

This book has become for many an indispensable text for the teaching of macroeconomics, valued for its monetarist approach and its lucid and extensive use of computer-simulated models as a teaching device. "A 'classic' in the sense that every serious student of macroeconomics is likely to want it in his or her library."—John P. Gould, University of Chicago

$14.95 384 pages (est.)

MISREGULATING TELEVISION

Network Dominance and the FCC

Stanley M. Besen, Thomas G. Krattenmaker, A. Richard Metzger, Jr., *and* **John R. Woodbury**

"I wish sleepless nights on all policymakers who fail to read the brief last chapter."
—Christopher Podmore, *Telecommunications Policy*

$9.95 210 pages

The University of CHICAGO Press
5801 South Ellis Avenue, Chicago, IL 60637

David Friedman's

PRICE THEORY...

"David Friedman's ***Price Theory*** *will, I predict, become familiar to students through lectures of their instructors, who will find the book extraordinarily interesting and instructive lecture material. The book provides interesting applications the students can grasp while reading and contemplating, permitting the instructor to concentrate on clarification of principles, concepts and diagrammatic technique, for which the classroom is better suited."*

—Arman Alchian

"This unorthodox new textbook doesn't just teach basic economics; it celebrates the discipline as a sort of cosmic nutpick, a key to everything from child rearing to conventional warfare."

—Jolie Solomon
The Wall Street Journal

"I highly recommend this text to students and others with curiosity about... our world. David Friedman obviously enjoys his work, for his enthusiasm is manifest on almost every page."

—Gary S. Becker

" This space reserved for you. "
Send for your exam copy today.

SOUTH-WESTERN
COLLEGE DIVISION

5101 Madison Road, Cincinnati, OH 45227 (513) 271-8811

AEA sponsored Group Life Insurance for you and your family—at attractive rates!

The AEA Group Life Insurance Plan can help provide valuable supplementary protection—at attractive rates—for eligible members and their dependents.

Because AEA participates in a large Insurance Trust which includes other scientific and technical organizations, the low cost may be even further reduced by premium credits. In the past nine years, insured members received credits on their April 1 semiannual payment notices averaging over 40% of their annual premium contributions. (These credits are based on the amount paid during the previous policy year ending September 30.) Of course future premium credits, and their amounts, cannot be promised or guaranteed.

Now may be a good time for you to re-evaluate your present coverage and look into AEA Life Insurance. Just fill out and return the coupon for more details at no obligation.

Administrator, AEA Group Insurance Program
1255 23rd Street, N.W.
Washington, D.C. 20037

H-8

Please send me more information about the AEA Life Insurance Plan.

Name ______________________ Age ________

Address ______________________________

City ________________ State ________ Zip ________

Or—call today Toll-Free 800-424-9883
(Washington, DC area, call 296-8030)

Ford Foundation Fellowships in European Society and Western Security

Harvard University's Center for International Affairs and Center for European Studies, in collaboration with the Ford Foundation, announce dissertation and post-doctoral support for research on the relationship between European society and Western security. This program places special emphasis on major policy issues and alternatives facing Europeans in the field of security, broadly defined, and the internal factors that influence European choices among these alternatives. The centers are particularly interested in building bridges among the social sciences and between European studies and strategic/defense studies. Past grants have been awarded for work in economics, history, political science, and sociology, and to Fellows from ten colleges and universities nationwide. Applications from women and minorities are especially welcome. Fellows selected will spend the 1987-88 academic year at the Center for International Affairs and the Center for European Studies pursuing their studies in the Fellowship topic area and participating in a research seminar. The Dissertation Fellowship carries a stipend of $11,000 and the Post-Doctoral Fellowship, a stipend of $22,000. The deadline for applications is February 17, 1987, and committee will reach a decision by April 1, 1987. For more information, write: Fellowship Office, Room 402, Center for International Affairs, Harvard University, 1737 Cambridge St., Cambridge, MA 02138; or call (617)495-1669.

St. Martin's 1987 economics titles... addressing the major issues

---

International Economic Problems and Policies
H. PETER GRAY, Rutgers University

A nontechnical, issue-oriented introduction to international economic phenomena and their influence on domestic policies, this book examines both the macro issues of commercial policy (trade) and the mechanics of international payments (finance). Gray provides unique in-depth coverage of multinational corporations, focusing on such important real-world issues as the use and misuse of trade warfare and the effects of international trade on employment.
496 pages (probable).
Publication: February 1987
Instructor's Manual available

Urban Economics and Public Policy
Third Edition
JAMES HEILBRUN, Fordham University

A clear, comprehensive introduction to the study of urban economics, this text offers a balanced treatment of urban economic theory and public policy issues. The book first covers growth theory, location theory, and the system of cities; and then applies these theories to such key urban problem areas as transportation, poverty, housing, and public finance. This new edition is updated with current statistics, and the policy chapters are fully revised to reflect changes introduced during the Reagan administration.
554 pages (probable)
Publication: February 1987

To request an examination copy of either of these titles, please write us on your college letterhead specifying your course title, present text, and approximate enrollment. Send your request to:

ST. MARTIN'S PRESS

Department JR, 175 Fifth Avenue, New York, New York 10010 SMP

NORTH-HOLLAND
NEW BOOKS IN ECONOMICS

Contributions to Mathematical Economics
In Honor of Gérard Debreu
W. Hildenbrand and **A. Mas-Colell** *(editors)*
US $69.00
ISBN 0-444-87924-2

International Trade and Exchange Rates in the Late Eighties
Proc. of the Conference
Namur/Brussels/Leuven, Belgium,
6–8 June 1985
T. Peeters, P. Praet and P. Reding *(editors)*
US $80.00
ISBN 0-444-87941-2

Public Enterprise Economics
Theory and Application
D. Bös
Advanced Textbooks in Economics, 23
US $32.50
ISBN 0-444-87899-8

Optimal Control Theory with Economic Applications
A. Selerstad and K. Sydsæter
Advanced Textbooks in Economics, 24
US $49.50
ISBN 0-444-87923-4

Capital Markets and Prices: Valuing Uncertain Income Streams
C.G. Krouse
Advanced Textbooks in Economics, 25
US $49.00
ISBN 0-444-87931-5

Rivalrous Consonance: A Theory of General Oligopolistic Equilibrium
R.E. Kuenne
Contributions to Economic Analysis, 157
US $30.00
ISBN 0-444-87937-4

The Dynamics of Market Economies
R.H. Day and G. Eliasson *(editors)*
US $65.00
ISBN 0-444-87841-6

International Trade: Surveys of Theory and Policy
Selections from the Handbook of International Economics
R.W. Jones *(editor)*
US $29.75
ISBN 0-444-87950-1

Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti
P.K. Goel and A. Zellner *(editor)*
Studies in Bayesian Econometrics and Statistics, 6
US $62.50
ISBN 0-444-87712-6

Improved Methods of Inference in Econometrics
G. Judge and T.A. Yancey
Studies in Mathematical and Managerial Economics, 34
US $59.00
ISBN 0-444-87936-6

NORTH-HOLLAND

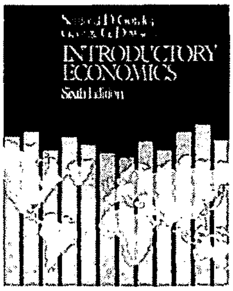
**ELSEVIER SCIENCE
PUBLISHING CO., INC.
P.O. BOX 1663
GRAND CENTRAL STATION
NEW YORK, NY 10163, USA**

US $ prices are valid only in the USA and Canada. In all other countries the Dutch Guilder price is definitive. Customers in the Netherlands, please add 5% B.T.W. In New York State applicable sales tax should be added. All prices are subject to change without prior notice.

NH/ECON/BK/0949

THE NEW EDITION OF A PROVEN SUCCESS FOR YOUR ONE-TERM COURSE

INTRODUCTORY ECONOMICS
Sixth Edition
Sanford D. Gordon, Russell Sage College,
New York State Council on Economic Education
George G. Dawson, Empire State College
Center for Business and Economic Education,
State University of New York
1987 Cloth 490 pages
1987 Paper 490 pages
*Instructor's Guide / Test Item File / Study
Guide / Archive Computerized Test Program
(for Apple® and IBM PC®)



Practical, real-world problems are the focus of this lively and accessible introductory economics text, with topics in international economics woven into every chapter. Based on an analytical/problem-solving approach, the book's proven teaching and learning system features four broad review models — demand and supply, long-run equilibrium, pricing factors of production, and national income determination. Within this conceptual framework, students are introduced to microeconomic and macroeconomic theory with applications to current problems and issues. New material has been added on federal tax reform, the value-added tax, employee buyouts, indicative planning in France and Japan, and recent changes in the Chinese economy. Sections on deregulation and the theory of rational expectations have been expanded, and a simple graphic summary explains the reasons behind the U.S. agricultural depression of the mid-1980's, providing an excellent review of demand and supply. Throughout, more emphasis is given to the special economic problems of consumers, women, minorities, and the "new poor."

Highlights:

- A new emphasis in the Sixth Edition on international economics reflects the growing importance to Americans of the world economy.
- A new feature, "Controversies in Our Times," appears in many chapters to highlight such controversial issues in economics as mergers and acquisitions, nuclear energy, and the public debt.
- Boxed items highlight current events which illustrate major principles discussed in the chapters.
- Charts, graphs, and statistics have been updated throughout the text.
- Highly successful learning aids include: chapter objectives, summaries, highlights, key terms and names, marginal notes to reinforce concepts, review questions and problems, boxed items to identify and explain common mistakes, and a greatly expanded glossary of terms.

*Apple is a registered trademark of Apple Computer, Inc.
**IBM PC is a registered trademark of International Business Machines Corp.

For details or sample copies, **call us toll free at: (800) 235-3565.**

D.C. Heath and Company
College Division / 125 Spring Street / Lexington, MA 02173

New York University Press

Wealth and Poverty
The Economics of Distribution
GORDON TULLOCK
272 pages *$40.00*
0-8147-8174-8

Liberty, Market, and the State
Political Economy in the 1980s
JAMES M. BUCHANAN
320 pages *$45.00*
0-8147-1085-9

Game Theory for the Social Sciences
Second and Revised Edition
HERVÉ MOULIN
256 pages, paper *$18.00*
0-8147-5431-7
256 pages, cloth *$45.00*
0-8147-5430-9

Eighty-Nine Exercises with Solutions from Game Theory for the Social Sciences
HERVÉ MOULIN
156 pages, paper *$7.50*
0-8147-5433-3
156 pages, cloth *$25.00*
0-8147-5432-5

Agricultural Policy and Trade
Adjusting Domestic Programs in an International Framework
D. GALE JOHNSON, KENZO HEMMI, and PIERRE LARDINOIS
114 pages *$18.50*
0-8147-4167-3

The Theory of Duality and International Trade
PASQUALE M. SGRO
320 pages *$35.00*
0-8147-7864-X

The Logic of Economic Discovery
Neoclassical Economics and the Marginal Revolution
ROBERT M. FISHER
240 pages *$40.00*
0-8147-2581-3

Economic History and the History of Economics
MARK BLAUG
336 pages *$40.00*
0-8147-1090-5

NEW YORK UNIVERSITY PRESS

New York University Press

Economic Organization
Firms, Markets, and Policy Controls
OLIVER E. WILLIAMSON
288 pages *$45.00*
0-8147-9208-1

International Trade
Theoretical Issues
BHARAT R. HAZARI
320 pages *$47.50*
0-8147-3438-3

Public Financing in a Democratic Society
Vol. I : Social Goods, Taxation and Fiscal Policy
Vol. II: Fiscal Doctrine, Growth, and Institutions
RICHARD A. MUSGRAVE
Vol. I: 416 pages *$65.00*
0-8147-5428-7
Vol. II: 416 pages *$65.00*
0-8147-5429-5
The Set: *$120.00*
0-PUBV-OLSE-T

Subjectivism, Intelligibility and Economic Understanding
ISRAEL M. KIRZNER, Editor
336 pages *$35.00*
0-8147-4589-X

Human Desire and Economic Satisfaction
Essays on the Frontiers of Economics
TIBOR SCITOVSKY
288 pages *$45.00*
0-8147-7862-3

Arrow and the Ascent of Modern Economic Theory
GEORGE R. FEIWEL, Editor
768 pages *$75.00*
0-8147-2582-1

Arrow and the Foundations of the Theory of Economic Policy
GEORGE R. FEIWEL, Editor
832 pages *$75.00*
0-8147-2583-X

Controversies in Political Economy
Selected Essays by G.C. Harcourt
O. M. HAMOUDA, Editor
272 pages *$40.00*
0-8147-3437-5

WASHINGTON SQUARE, NEW YORK, N.Y. 10003
Or call: (914) 591-9111

BASIC Issues

New

Bailout
An Insider's Account of
Bank Failures and Rescues
IRVINE H. SPRAGUE

The Economic Emergence of Women
BARBARA R. BERGMANN

King Solomon's Mines Revisited
Western Interests and the
Burdened History of Southern Africa
WILLIAM MINTER

The Capitalist Revolution
Fifty Propostions About
Prosperity, Equality, and Liberty
PETER L. BERGER

The International Money Game
Fifth Edition, Revised
ROBERT Z. ALIBER

The Transformation of American Industrial Relations
THOMAS A. KOCHAN, HARRY C. KATZ,
& ROBERT B. McKERSIE

Forthcoming in Spring 1987

Manufacturing Matters
The Myth of the Post-Industrial Economy
STEPHEN S. COHEN & JOHN ZYSMAN
February

Social Security
The System that Works
MERTON C. BERNSTEIN
& JOAN BRODSHAUG BERNSTEIN
February

Running American Business
Top CEOs Rethink Their Major Decisions
ROBERT BOYDEN LAMB
April

Unheard Voices
Labor and Economic Policy
in a Competitive World
RAY MARSHALL
March

World Resources 1987
WORLD RESOURCES INSTITUTE &
THE INTERNATIONAL INSTITUTE FOR ENVIRONMENT AND DEVELOPMENT
March

Now in Paperback

CEO
Corporate Leadership in Action
HARRY LEVINSON & STUART ROSENTHAL

The Second Industrial Divide
Possibilities for Prosperity
MICHAEL J. PIORE & CHARLES F. SABEL

Forthcoming in Paperback

Beyond Human Scale
The Large Corporation at Risk
ELI GINZBERG & GEORGE VOJTA
April

Thinking Economically
How Economic Principles Can
Contribute to Clear Thinking
MAURICE LEVI
April

Basic Books, Inc.
10 East 53rd St., New York, NY 10022

Two new second editions from
Harlan Davidson, Inc.

Contemporary Developments in Financial Institutions and Markets, Second Edition

Edited by Thomas M. Havrilesky, *Duke University*, and Robert Schweitzer, *University of Delaware*

The fast pace of change in the financial institutions and markets area has made a major revision of this popular collection of readings absolutely necessary. Fully half of the original articles have been replaced in this edition by new readings on changes in the thrift industry; the repurchase agreement markets; securitization; futures markets; options markets; and interest rate swaps.

Like the first edition, this revision of **Contemporary Developments in Financial Institutions and Markets** contains readings by prominent financial economists that enable the student and general reader to understand the financial system and to grasp the mechanics and significance of recent major changes.

Contemporary Developments in Financial Institutions and Markets, Second Edition, is suitable for courses in Financial Institutions and in Money and Banking and related areas.

480 pages (t) paper

Introduction to Modern Macroeconomics, Second Edition

by Thomas M. Havrilesky, *Duke University*

Introduction to Modern Macroeconomics continues to gain ground as a successful alternative to both New Classical textbooks that many students find too difficult and texts using the obsolete LM-IS approach.

Introduction to Modern Macroeconomics, Second Edition, is accompanied by a complete study guide and a computer tutorial disk designed for use on IBM PC–compatible systems. The ancillary materials now help to make the theories and applications of intermediate macroeconomics understandable to an even wider range of students. But the text, carefully revised and improved throughout, stands alone as the means by which to bring the student to the frontiers of applied macroeconomics without requiring a technical background beyond the principles course.

450 pages (t) cloth edition
Student's Assistance Manual, 208 pages, paper edition

Examination copies of the texts and *Student's Assistance Manual* available on a 60-day approval basis only. Examination copies of the computer tutorial disk are not available. A master copy is free upon request with adoptions of the text.

Harlan Davidson, Inc.
3110 North Arlington Heights Road
Arlington Heights, Illinois 60004-1592

GENERAL EQUILIBRIUM MODEL FOR MICRO-COMPUTERS

GEMODEL 2.0

Solves open or closed, two or three-industry small open economy models employing two or three factors. Industries can have constant or diminishing returns to scale. Admits one to nineteen household groups with or without income/leisure choice. Solves with a wide array of factor, commodity and income taxes.

PRICE: US$ 395.00

Ideal to enrich
courses in Price
Theory, Trade and
Finance by solving
models
within seconds
at 48 different
levels of complexity
with
18 to 176 equations.

GESTATS

Calibrates GEMODEL parameters to data and elasticity assumptions. Checks data consistency.

PRICE: US$ 195.00*

GEREPORT

Compares GEMODEL simulation results with the base case or with other simulation results. Computes measures of welfare change and marginal costs of taxes.

PRICE: US$ 95.00*

GEDATA

18 data files for exercises in price theory, trade and finance using GEMODEL.

PRICE: US$ 25.00*

---

*When ordering, please specify whether it is for use with GEMODEL 1.7 or GEMODEL 2.0

SYSTEM REQUIREMENTS: All GEMODEL software requires only 256K, 2 drives, and DOS 2.1 (or better) in an IBM PC, XT, AT or compatible.

MATH CO-PROCESSOR Support: US$ 55.00 extra for any or all programs supplied on one diskette.

DEMONSTRATOR diskette for GEMODEL 2.0: US$ 25.00

Users of earlier versions of GEMODEL can order the GEMODEL 2.0 update for only US$ 95.00.

DISCOUNTS: For pre-paid order of GEMODEL, Gestats, GEREPORT and GEDATA on one purchase order: 10% off list price.
10 copies of any of the above software on one order: 70% off list price.
20 copies of any of the above software on one order: 80% off list price.

TO ORDER write to:

D.I.A. Agency Inc., 1879 Kingsdale Ave., Ottawa, Ontario, CANADA K1T 1H9

Ask also about our **ECONOMISTS' TOOLKIT**, a selection of software for the economist.

DIA

BALLINGER

AMERICAN SOCIETY

Public and Private Responsibilities

Edited by **Winthrop Knowlton** and **Richard Zeckhauser**

Prepared under the auspices of the Center for Business and Government in celebration of Harvard University's 350th and the John F. Kennedy School's 50th anniversaries.

" . . .an excellent volume of essays by distinguished authors on the basic question for the structure of American government: the allocation of responsibilities between the public and private sectors.... The essays are filled with wisdom, balance, and occasional wit."
— Kenneth J. Arrow, Stanford University
Nobel Laureate in Economics, 1972

"**American Society's** essays . . . are challenging, wise, thought-provoking and must reading for the players in both sectors."
— Irving S. Shapiro
former Chairman of the Board, Dupont
Partner, Skadden, Arps, Slate, Meagher, & Flom

1986 0-88730-129-0 $24.95, cloth 360 pages

TECHNOLOGY AND ECONOMIC POLICY

Edited by **Ralph Landau** and **Dale W. Jorgenson**

This book fills an important gap in the debate over economic policy by providing a technological perspective on business taxation, the federal deficit, the cost of capital, and the international competitiveness of the U.S. economy. The eminent contributors offer a thorough discussion of U.S. economic policy, particularly as it affects technology, and provide a realistic agenda for future research.

1986 0-88730-068-5 $29.95, cloth 376 pages

MONETARISM AND LIBERALIZATION

The Chilean Experiment

Sebastian Edwards and **Alejandra C. Edwards**

An in-depth examination of the Pinochet government's experiment in free-market policies from promising beginnings in 1973 to ultimate failure in 1982, **Monetarism and Liberalization** looks at what went wrong in Chile. Why did liberalization and stabilization programs initially succeed in dramatically reducing inflation and raising real output only to collapse nine years later? This analysis illuminates some of the more puzzling aspects of the Chilean experiment and extracts vital lessons that will be helpful to economists and policymakers.

1987 0-88730-105-3 $26.95, cloth 240 pages

THE ECONOMICS OF TELECOMMUNICATIONS

Theory and Policy

John T. Wenders

With the breakup of AT&T and the onset of "telephone wars" between AT&T and companies like MCI and Sprint, telephone service as we knew it changed for good. In **The Economics of Telecommunications**, John Wenders, a consultant to AT&T before its divestiture, explains the economic and political forces behind the breakup and looks at the future of the telecommunications industry. He develops a framework for understanding the economic basis for regulation of the industry and then argues that deregulation will be beneficial. Wenders calls for rapid changes in the industry, including the raising of local calling charges, the lowering of toll calling rates, and the implementation of local measured service.

1987 0-88730-119-3 $34.95, cloth 296 pages

BALLINGER

YES! Please send me:
____**American Society** @ $24.95 (6610919)
____**Technology and Economic Policy** @ $29.95 (6610331)
____**Monetarism and Liberalization** @ $26.95 (6610729)
____**Economics of Telecommunications** @ $34.95 (6610844)
☐ Payment enclosed ☐ Bill me
Charge my ☐ MC ☐ VISA ☐ AMX
Card no. ____________ Exp. date______
Signature____________
Send to: ____________
____
____________Zip______

Harper & Row
Order Department
2350 Virginia Avenue, Hagerstown, MD 21740
(800) 638-3030

My state sales tax $______
Postage/handling ($1.50/bk)* $______
***Prepaid orders are postage free!**
TOTAL $______

Prices subject to change. All orders subject to credit approval. U.S. funds only. If you order by phone, tell the operator your order code is **AAER1286**

THE ECONOMICS INSTITUTE

Gateway to Successful
Master's and Doctoral Degree Studies
in Economics, Agricultural Economics, Business and Administration
for STUDENTS FROM ABROAD

Since its establishment in 1958 under the sponsorship of the American Economic Association, the Institute has provided specialized preparatory training and orientation to over 6,000 students from over 120 countries en route to over 320 universities in the United States and other English-speaking countries.

PROGRAM FEATURES

- Five to forty-five week individualized programs.
- Intensive and comprehensive review or supplementary preparatory training in:
  English, Computer Usage, Economic Theory, Mathematics, Statistics, Accounting, Management, Finance and Marketing.
- Equivalent courses and advisory reports facilitate integration with subsequent university programs.
- Standardized test preparation (TOEFL, GRE, GMAT).
- Orientation to U.S. campus and community life.
- University placement assistance services.
- Certificate and Diploma Programs for short-term professional trainees.

END RESULTS for

- **Associated universities:** Improved admissions procedures and a source of additional high-quality foreign students.
- **Foreign students:** Greater accessibility to U.S. universities.
- **Sponsoring organizations:** Reduced total costs.
- **And for all three:** Better grade performance records in degree programs and more rapid completion of degree requirements.

ORGANIZATION

AEA Policy and Advisory Board
Edwin S. Mills, Princeton University, Chairman
Lance E. Davis, California Institute of Technology
W. Lee Hansen, University of Wisconsin, Madison
Joseph Havlicek, Ohio State University
Teh-wei Hu, Pennsylvania State University
Samuel A. Morley, Vanderbilt University
Dwight H. Perkins, Harvard University
Stefan H. Robock, Columbia University

Director
Wyn F. Owen, University of Colorado

The Economics Institute, 1030 13th Street, Boulder, Colorado, 80302, USA;
Telephone: (303) 492-3000; Telex: 450385 ECONINST BDR

NEW FROM PENNSYLVANIA

Economic Development Within the Philadelphia Metropolitan Area

ANITA A. SUMMERS and
THOMAS F. LUCE

The 1986 report on *Economic Development Within the Philadelphia Metropolitan Area* focuses on economic relationships within the eight counties in the Philadelphia Primary Metropolitan Statistical Area (PMSA). It complements the 1985 report on the region as a whole.

Following an update of last year's report, which documents recent improvements in the regional economy, the 1986 report analyzes the decentralized metropolitan map. Long-term economic developments, changes in socio-economic profiles between 1960 and 1980, and patterns of employment are examined on a county-by-county basis.

Each county's industrial strengths are identified. Special attention is given to the spread and growth of employment in high-technology industries, the interdependencies between jobs and residents in the city and suburbs, and the role of federal and state aid to the region.

The 1986 report on *Economic Development Within the Philadelphia Metropolitan Area* is of value to scholars; government, business, and labor leaders; local civic organizations; and the news media.

180 pages. Tables, graphs, statistical appendices, CIP. Available in December 1986. Paper, ISBN 0-8122-1231-2, $12.95.

Economic Report on the Philadelphia Metropolitan Area: 1985
Anita A. Summers and Thomas Luce

The first report accomplishes three objectives. It assembles data relevant to the analysis of the region's potential for economic growth, provides a detailed analysis of the dynamic patterns of the area's industry, and develops a list of options for responding to these patterns.
144 pages. Paper, ISBN 0-8122-1212-6, $9.95

upp UNIVERSITY OF PENNSYLVANIA PRESS
Blockley Hall / 418 Service Drive / Philadelphia, PA 19104-6097

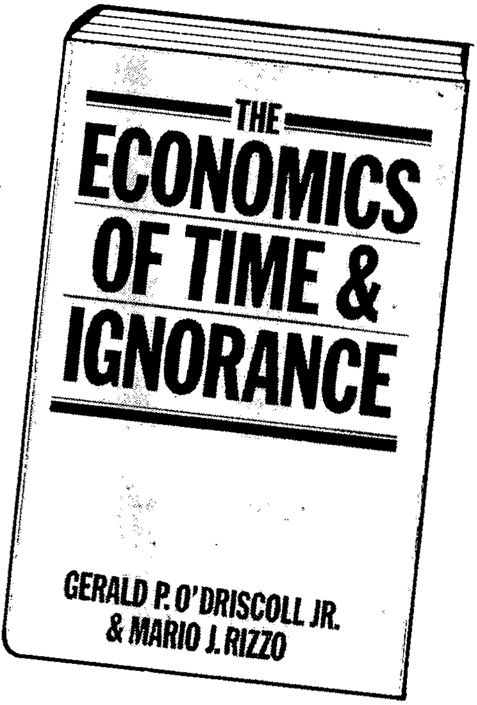
Basil Blackwell

*"A superb book."**

Now Available in Paper

*"In terms of style, tone, and execution, it is flawless ... it is an original, and indeed a very creative work. The book is concisely written, and warrants a close reading. But it is also well-written.... **The Economics of Time and Ignorance** has all the makings of a classic.... [T]his is the sort of book that every economist should read. No matter what one's prior beliefs about economics are, this book will provoke thought."*
—Bruce J. Caldwell,
*Southern Economic Journal**

"The authors' discussions ... all hit home.... Perhaps now the Austrians will be allowed to help renovate economics."
—*Times Literary Supplement*



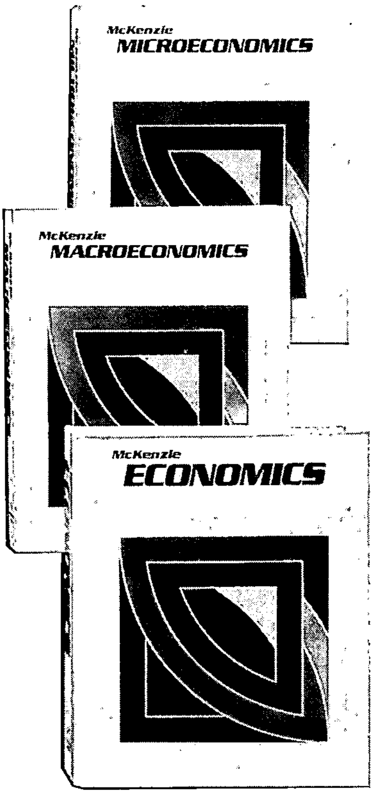
This book is the first contemporary account of the foundations of Austrian subjectivist economics. O'Driscoll and Rizzo, two of the school's leading modern-day exponents, present an integrated view of its themes, making an original contribution to our understanding of uncertainty and dynamic processes. In so doing, they reinstate individual choice to its rightful place at the center of the economic system.

$14.95 paper 261 pp. 0-631-15484-1
[34.95 cloth 0-631-13343-7]

Toll-free Ordering: 1-800-638-3030

Basil Blackwell
432 Park Avenue South • Suite 1503 • New York, NY 10016

BENEFIT FROM THE EXPERIENCE



RICHARD McKENZIE
Clemson University

ECONOMICS
Complete Hardcover Edition: 847 pages
MACROECONOMICS: paper • 477 pages
MICROECONOMICS: paper • 569 pages
Study Guide • Instructor's Manual
Test Bank • Computerized Test Bank
GPA: Grade Performance Analyzer
Transparencies • Computer Graphics
Package • 1986

Combining experience in national policy-making and in teaching, Richard McKenzie offers a textbook that presents economic principles in terms everyone can understand.

For adoption consideration, request an examination package from your regional Houghton Mifflin office.

Houghton Mifflin Company
13400 Midway Rd., Dallas, TX 75244-5165
1900 S. Batavia Ave., Geneva, IL 60134
989 Lenox Dr., Lawrenceville, NJ 08648
777 California Ave., Palo Alto, CA 94304

JOB OPENINGS FOR ECONOMISTS

Available only to AEA members and institutions that agree to list their openings.

Annual Subscription Rates

U.S.A., Canada, and Mexico (first class):	$15.00, regular AEA members and institutions $ 7.50, junior members of AEA
All other countries (air mail):	$22.50, regular AEA members and institutions $15.00, junior members of AEA

Please begin my issues with:
☐ February ☐ April ☐ June ☐ August ☐ October ☐ December

Name____
First Middle Last

Address_______________________________________

____
City State/Country Zip/PostalCode

Check one:
☐ I am a member of the American Economic Association.
☐ I would like to become a member. My application and payment are enclosed.
☐ (For institutions) We agree to list our vacancies in JOE.

Send payment (U.S. currency only) to:

THE AMERICAN ECONOMIC ASSOCIATION
1313 21st Avenue South
Nashville, Tennessee 37212

NOW AVAILABLE . . .
CAN-AM STATISTICAL PACKAGE II

CAN-AM STATISTICAL PACKAGE II features:

- **Descriptive Statistics** —
  - Mean, variance & stand deviation
  - Linear regression & correlation
  - Analysis of variance
  - Spearman's rank correlation
  - Mann-Whitney U test
- **Time Series** — Moving average; exponential smoothing
- **Matrix Operation** — Matrix multiplication & inversion
- **Regression Analysis** — Ordinary least squares (OLS)
  - Polynomial distributed lag (PDL)
  - Two-stage least squares (TSLS)

- Cochrane-Orcutt estimation feature in OLS, PDL and TSLS
- More than usual summary statistics including autocorrelation coeff., D.W., log of likelihood, corr matrix, var-cov matrix for estimated coeff., etc.
- Prediction with confidence intervals in OLS
- Handles large numbers of variables & observations; variable transformation
- Prices in U.S. dollars: $99.90 for IBM (PC, XT, jr.); $79.90 for Apple (II+, IIe, IIc); $59.90 for Commodore (64, 128); and $12.00 for demo disk (non-refundable but to be subtracted from the purchase price)
- To order: send cheque, money order or credit card number with expiry date (Visa or MasterCard only) to

CAN-AM FINANCIAL CONSULTING
177 Caddy Ave., Sault St. Marie, Ont. P6A 6H7 CANADA

Happy Holidays!

from
McGraw-Hill and
Campbell McConnell

Study Guide
Test Bank II
Test Bank I
Intructor's Resource Manual
Overhead Transparencies to Accompany ECONOMICS
ECONOMICS
Transparency Masters
ECONOMICS
DRI Data Diskette McConnell's ECONOMICS 10/e
Interactive Economic Graphic Software McConnell's ECONOMICS 10/e
Interactive Economic Graphic Software McConnell's ECONOMICS 10/e
TENTH EDITION

McGraw-Hill Book Company
College Division
1221 Avenue of the Americas
New York, New York 10020

New and Noteworthy

U.S. Trade Policies in a Changing World Economy

edited by Robert H. Stern

"A timely volume by leading experts as the U.S. is abandoning its commitment to the postwar multilateral trading system and U.S. trade policy is drifting away from the principle of non-discrimination."—Jaime de Melo, Senior Economist, The World Bank
$25.00

Superfairness

Applications and Theory

William J. Baumol

"William Baumol's work gives economists a theory of equity that can be applied with the rigor, and taught with the clarity, of Paretian efficiency theory. Everyone who teaches or uses welfare economics will benefit from reading this book."—Duncan K. Foley, Professor of Economics, Barnard College of Columbia University
$20.00

Microtheory

Applications and Origins

William J. Baumol

These essays provide an engaging intellectual history of one of the leading figures in the field of economics.
$35.00

Dollars, Debts, and Deficits

Rüdiger Dornbusch

"Professor Dornbusch provides a clear, fascinating, timely and provocative analysis of the key political issues faced by economists and policy makers concerned with the world economy."
—Jacob A. Frenkel, David Rockefeller Professor of International Economics, University of Chicago
$20.00

Perspectives on Safe and Sound Banking

Past, Present, and Future

George J. Benston, Robert A. Eisenbeis, Paul M. Horvitz, Edward J. Kane, and George G. Kaufman

Copublished with the American Banking Association.
$19.95

The MIT Press

28 Carleton Street, Cambridge, MA 02142

Who's Who in Economics

A Biographical Dictionary of Major Economists 1700–1984

Second Edition *edited by Mark Blaug*

"A volume of real usefulness . . . exceedingly useful in our day to day work."—*History of Political Economy.*

This new edition includes 337 more entries, bringing the total to over 1,400 expanded sketches of prominent economists of the past three centuries.

$95.00

The MIT Dictionary of Modern Economics

Third Edition *edited by David W. Pearce*

The MIT Dictionary of Modern Economics provides an authoritative reference guide to economic terminology and nomenclature. This new edition adds 150 definitions and revises many others, with a total of more than 2,650 entries.

$12.50 paperback (Cloth, $35.00)

Japanese Financial Markets

Deficits, Dilemmas, and Deregulation

Robert Alan Feldman

"Dr. Feldman's new book describes the essential features of Japan's financial innovation, deregulation and reform with international perspective on the basis of penetrating analysis of their causes and consequences."—Yoshio Suzuki, The Bank of Japan

$27.50

The Invisible Link

Japan's Sogo Shosha and the Organization of Trade

M. Y. Yoshino and Thomas B. Lifson

"This book puts the sogo shosha under the microscope. It examines the specifics of how they coordinate trade and the flow of goods through the production process and how management keeps these far-flung enterprises working together despite internal conflicts of interest. Along the way, it offers perhaps the best description available of the internal dynamics of a Japanese office."—*Business Week*

$19.95

Now in paperback

The Age of Interdependence

Economic Policy in a Shrinking World

Michael Stewart

"Not only economists should be interested in this well-conceived (and well-written) book which grapples with two basic weaknesses of conventional macroeconomic policy, Keynesian or monetarist."—*Foreign Affairs*

$7.95 paper

Order today, Mastercard and Visa accepted—(617) 253-2884.

INSTITUTE FOR INTERNATIONAL ECONOMICS

New and Recent Books

11 Dupont Circle, NW,
Washington, DC 20036 (202) 328-9000

SYSTEM UNDER STRESS: *New*
THE POLITICS OF UNITED STATES TRADE POLICY
I.M. Destler

Describes the postwar American system for combating protectionist pressures and how this system has eroded over the past 15 years, compounded by pressures from the unprecedented US trade deficit of the 1980s. Concludes with proposals aimed at restoring the nation's capacity to pursue rational trade policies. Published jointly with The Twentieth Century Fund.
November 1986/384 pp./ISBN 0-88132-058-7/
$30.00 cloth/057-9/$18.00 paper

THE FUTURE OF WORLD TRADE IN TEXTILES *New*
William R. Cline

Analyzes the economics of the textile industry and the impact of macroeconomic forces on it, examines the costs and benefits of existing and proposed controls for different countries and groups, and makes proposals for the future conduct of trade in textiles and apparel.
POLICY ANALYSES IN INTERNATIONAL ECONOMICS Series
December 1986/c. 180 pp./
ISBN 0-88132-025-0/$10.00 paper

THE UNITED STATES-JAPAN ECONOMIC PROBLEM
C. Fred Bergsten and William R. Cline

"Absolutely first-rate, convincing and cogent. . . . A truly remarkable and important study." —Peter F. Drucker
"[It] provides a sound list of policy suggestions as well as a lucid, and thus educational, analysis of the current situation." —Gary N. Horlick, *World Economy*
POLICY ANALYSES IN INTERNATIONAL ECONOMICS 13
1985/174 pp./ISBN 0-88132-039-0/$10.00 paper

TRADE PROTECTION IN THE UNITED STATES:
31 CASE STUDIES
Gary Clyde Hufbauer, Diane T. Berliner, and Kimberly Ann Elliott

Case studies of 31 major US industries that have received some form of special protection. Each case contains a brief summary of relief action, estimates of the costs to consumers, the gains to producers and foreigners, and the overall efficiency and welfare losses to the US economy.
March 1986/371 pp./
ISBN 0-88132-040-4/$25.00 paper

To order, or for complete catalog, write Dept. AR:
Institute for International Economics,
11 Dupont Circle, NW, Washington, DC 20036

UNITED STATES-CANADIAN *New*
INTERDEPENDENCE: THE QUEST FOR FREER TRADE
Paul Wonnacott

Analyzes the potential impact of new trade arrangements between the United States and Canada and the difficult problems that must be addressed in coming negotiations—including quite different perceptions as to what constitutes fair trade.
POLICY ANALYSES IN INTERNATIONAL ECONOMICS Series
December 1986/c. 100 pp./
ISBN 0-88132-056-0/$10.00 paper

AGRICULTURE AND THE GATT: *New*
ISSUES IN A NEW TRADE ROUND
Dale E. Hathaway

Analyzes the nature and extent of the farm trade problem, identifies the fundamental issues that must be addressed, and examines the interests of the major parties involved. Concludes with a series of practical policy recommendations for developing a better regime for agricultural trade.
POLICY ANALYSES IN INTERNATIONAL ECONOMICS Series
December 1986/c. 120 pp./
ISBN 0-88132-052-8/$10.00 paper

TOWARD RENEWED ECONOMIC GROWTH *New*
IN LATIN AMERICA
Bela Balassa, Gerardo M. Bueno, Pedro-Pablo Kuczynski, and Mario Henrique Simonsen

Proposes a comprehensive strategy for Latin America which could foster self-sustaining growth and simultaneously permit continued servicing of the external debt. Specific recommendations are made for policy changes to facilitate the needed reforms. Sponsored jointly by the Institute, El Colegio de Mexico, and the Getúlio Vargas Foundation of Brazil.
September 1986/207 pp./
ISBN 0-88132-045-5/$15.00 paper

DEFICITS AND THE DOLLAR:
THE WORLD ECONOMY AT RISK
Stephen Marris

"A first class scholarly study which is stimulating and elegantly written."
—Alan A. Tait, *Journal of Public Policy*
"Stands out for the thorough and sophisticated way it links the domestic features of major national economies with the international disequilibrium they have produced."
—William Diebold, Jr., *Foreign Affairs*
POLICY ANALYSES IN INTERNATIONAL ECONOMICS 14
1985/381 pp./ISBN 0-88132-030-7/$15.00 paper

ESSAY COMPETITION IN TRANSPORT FOR YOUNG RESEARCHERS

The Trustees of the *REES JEFFREYS ROAD FUND*, through the JOURNAL OF TRANSPORT ECONOMICS AND POLICY are offering a prize of £500 for an essay in transport.

The subject must be within the Journal's field of interest. The essay must report research which has not been published, or offered for publication, elsewhere.

Entrants must be under 30 years of age on the closing date for entries, which is *30 September 1987.*

Interested researchers should obtain further details from:

The Secretary,
Journal of Transport Economics and Policy,
University of Bath,
Claverton Down,
Bath BA2 7AY,
U. K.

Now more powerful than ever...MicroTSP™ 5.0

Now available in enhanced version 5.0, this software package for regression, forecasting, and time series analysis—loaded with exciting new features—gives you all the power you need to analyze time series data and work cross-section problems in any field.

For financial analysis...sales forecasting...cost analysis and forecasting...or any other type of time series forecasting or modeling...MicroTSP more than meets your professional needs.

Well-written documentation.
320 page User's Manual by Robert Hall, Stanford Univ. Included with each package: sample data and case studies, regular and 80(2)87 program disks.

...High performance software for statistics and forecasting

MicroTSP machine specifications:
Version 5/0: IBM/PC, XT, AT, and compatibles, 384K RAM, 2 disk drives.

NEW FEATURES:

- **Expanded memory capacity**—up to 32,000 data points
- **New techniques:** Nonlinear least squares, probit, logit
- **Improved hard disk support**—no key disk required
- **Hercules graphics support**

Order now by filling out and mailing the coupon below
or call TOLL-FREE 1-800-782-3737 (in New York 212-512-2345)

QMS Quantitative Micro Software

Clip and mail to: E.T. Matthews—26th Fl.
McGRAW-HILL BOOK COMPANY
1221 Avenue of the Americas
New York, NY 10124

I want to order MicroTSP. Please send me:	**Price**
☐ Information on trade-up policy for IBM/PC	
☐ IBM/PC (5.0) Demo Package 0-07-852207-2	$ 30.00
☐ IBM/PC (5.0) 0-07-852205-6	$595.00

IBM/PC, XT, and AT are registered trademarks of International Business Machines. MicroTSP is a registered trademark of Quantitative Micro Software.

PRICES AND TERMS VARY OUTSIDE U.S.—FOR THIS INFORMATION WRITE B. MUTINSKY AT ADDRESS ABOVE.

☐ My check (including local sales tax) is enclosed.
Charge my ☐VISA ☐MASTERCARD Exp. Date__________
Card #__________
Signature__________
Name__________
Address__________
City________ State________ Zip________

Payment must accompany order. Check payment to McGraw-Hill Book Co. (Include local sales tax). Disks are not returnable.

TSP-8

American Economic Association

Summer Minority Program

at Temple University

June 3, 1987—July 29, 1987

A New Economics Program with Exciting Features

An Attractive Stipend – plus Books,
Room, Board, Tuition and Travel are
Available to Qualified Students.

For More Information Contact:

THE HARRY A. COCHRAN RESEARCH CENTER
ROOM 111
SPEAKMAN HALL
TEMPLE UNIVERSITY
PHILADELPHIA, PA 19122

(215-787-6750)

SCHOLARLY INVESTMENTS

Tracing New Orbits
Cooperation and Competition in Global Satellite Development

Edited by Donna A. Demac

The first comprehensive overview of emerging international satellite activity and the issues raised by it. A broad range of experts discuss regulation, bilateral cooperation and competition, shared resources, and the potential to promote peace.
Columbia Studies in Business, Government, and Society
Eli M. Noam, General Editor
320 pp., $35.00

The Changing Structure of American Banking

Peter S. Rose

A timely and thorough study of bank structures and competition, financial services, and the effects of deregulation.
Columbia Studies in Business, Government, and Society
392 pp., $40.00

Neorealism and Its Critics

Edited by Robert O. Keohane

"An excellent guide through the thicket of controversies on some of the central questions of international politics.... some of the best discussions of Neorealism, its utility, its limits, and its defects."
—Robert Jervis
The Political Economy of International Change
John Gerard Ruggie, General Editor
256 pp., $13.00 pa, $45.00 cl

The New Order of the Oceans
The Advent of a Managed Environment

Edited by Giulio Pontecorvo
Foreword by Elliot L. Richardson

These articles place recent developments in ocean policy within the context of economic development and international relations.
240 pp., $32.50

Africa in Economic Crisis

Edited by John Ravenhill

"A thorough, sober, and balanced analysis. It contains critical essays on the roots, manifestations, and competing solutions offered to remedy Africa's current economic stagnation."—Naomi Chazan
320 pp., $13.00 pa, $35.00 cl

Workers on the Edge
Work, Leisure, and Politics in Industrializing Cincinnati, 1788–1890

Steven J. Ross

"This excellent book merits special attention.... Ross offers a beautifully researched, ironic account of local unionism, electoral politics, and working-class leisure from the Age of Jackson onward."
—*American Historical Review*
Columbia History of Urban Life Series
Kenneth T. Jackson, General Editor
432 pp., $35.00

Political Facts of the United States Since 1789

Erik W. Austin
with the assistance of Jerome M. Clubb

Brings together a wealth of statistics on American political history from the nation's origins up through recent elections, including sections on wealth, revenue, taxation, and public expenditure.
450 pp., $60.00

Dossier Society
Value Choices in the Design of National Information Systems

Kenneth C. Laudon

"A wonderfully penetrating, insightful, and well-written examination of one of the most troubling challenges now facing the American people."
—David Burnham, *The New York Times*
Computing, Organizations, Policy, and Society (CORPS)
Rob Kling and Kenneth L. Kraemer, General Editors
480 pp., $45.00

Now in paperback

Town in the Ruhr
A Social History of Bochum, 1860–1914

David F. Crew

"From now on even historians who have not the slightest interest in Bochum will need to reckon with this book."—*Journal of European Economic History*
289 pp., $13.00 pa

To order, send check or money order to Dept. JN at the address below, including $2.00 for postage and handling.

Columbia University Press
136 South Broadway, Irvington, NY 10533

SPREADSHEET STATISTICAL/FORECASTING PROGRAM ySTAT™

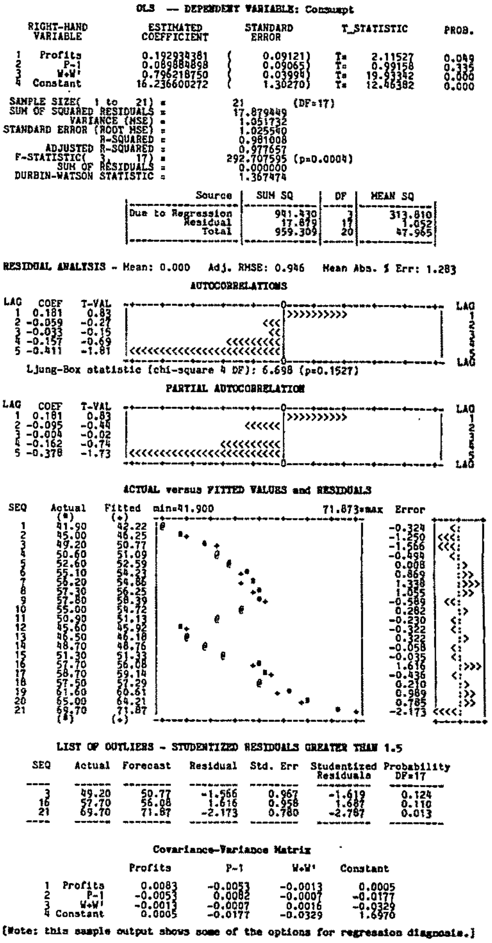
for Your IBM Personal Computer

WORKS LIKE LOTUS 1-2-3, WITH FEATURES LIKE SAS

- **ySTAT is the only full-fledged spreadsheet statistical program** - a fast, powerful, easy-to-use stand-alone program.
- **Menu-driven format as in Lotus 1-2-3-** offers many functions that are not available even in 1-2-3, such as vector formulas, lagged variables, dummy variables, and row-column transposition.
- **You can read, enter, edit, copy, move, and sort the data** as well as group and partition the data for analysis right on the spreadsheet. A total of 40 function keys for one-keystroke operations.
- **Ten statistical function keys:** to tally sums and means; subtract variable means or seasonal means; standardize the data; aggregate seasonal data; generate seasonal indicators; perform moving averages, linear trend forecasts, and so on.
- **Automatic recording feature** for programming macro keys, which with one keystroke, will replicate an entire sequence of analyses. You can save macros for future uses.
- **ySTAT reads Lotus' files** and any free-format or fixed-format textfile generated by other PC programs or mainframe. The data can be numeric, alphabetic, including missing values.
- **No disk swapping:** ySTAT is compactly written in C - the entire program resides on one disk.
- **Descriptive statistics,** frequency distribution, correlation, crosstabs, analysis of variance.
- **Multiple regression** (OLS), standardized regression, weighted least squares, two-stage least squares, polynomial regression, and Cochrane-Orcutt method. Including **regression diagnostics.** (see sample output at right)
- Pooling of time-series and cross-sectional data.
- **Two-sample Difference of means test, nonparametric tests** (Kolmogorov-Smirnov test, Wald-Wolfowitz runs test, Mann-Whitney test, Wilcoxon matched-pairs signed-ranks test) and Spearmans's and Kendall's **rank-order correlations.**
- **Time-series models:** autoregressive model, moving average model, and **Box-Jenkins model.**
- **Nonseasonal forecasts:** linear trend, quadratic trend, polynomial model, sinusoidal model, and simple, double and triple exponential smoothings.
- **Seasonal forecasts:** exponential smoothings with seasonal indicators or trigonometric functions; Winter's additive and multiplicative procedures.

SYSTEM REQUIREMENTS: IBM PC, XT, AT, or compatibles; a math co-processor is optional. Also available for Hewlett-Packard Portable Plus.

PRICE: $395. Including one free update.



```
                 OLS  -- DEPENDENT VARIABLE: Consumpt

  RIGHT-HAND        ESTIMATED        STANDARD        T_STATISTIC     PROB.
   VARIABLE        COEFFICIENT        ERROR

 1    Profits      0.192934381   (     0.09121)   T=     2.11527     0.049
 2        P-1      0.089884898   (     0.09065)   T=     0.99158     0.335
 3        W+W'     0.796218750   (     0.03994)   T=    19.93342     0.000
 4   Constant     16.236600272   (     1.30270)   T=    12.46382     0.000

   SAMPLE SIZE(  1 to   21) =          21        (DF=17)
  SUM OF SQUARED RESIDUALS =          17.879449
            VARIANCE (MSE) =           1.051732
  STANDARD ERROR (ROOT MSE) =          1.025540
                 R-SQUARED =           0.981008
        ADJUSTED R-SQUARED =           0.977657
    F-STATISTIC(   3,   17) =        292.707595 (p=0.0004)
         SUM OF RESIDUALS =            0.000000
  DURBIN-WATSON STATISTIC =            1.367474

                          Source |  SUM SQ  |  DF |  MEAN SQ  |
                 |---------------|----------|-----|-----------|
                 |Due to Regression|  941.430|    3|    313.810|
                 |       Residual|    17.879|   17|      1.052|
                 |          Total|   959.309|   20|     47.965|
                 |---------------|----------|-----|-----------|

RESIDUAL ANALYSIS - Mean: 0.000   Adj. RMSE: 0.946   Mean Abs. $ Err: 1.283

                         AUTOCORRELATIONS

LAG   COEF    T-VAL   +----+----+----+----+----0----+----+----+----+----+  LAG
 1  0.181     0.83    |                         |>>>>>>>>>>                |   1
 2 -0.059    -0.27    |                      <<<|                          |   2
 3 -0.033    -0.15    |                       <<|                          |   3
 4 -0.157    -0.69    |                <<<<<<<<<|                          |   4
 5 -0.411    -1.81    |<<<<<<<<<<<<<<<<<<<<<<<<<|                          |   5
                      +----+----+----+----+----0----+----+----+----+----+  LAG
   Ljung-Box statistic (chi-square 4 DF): 6.698 (p=0.1527)

                      PARTIAL AUTOCORRELATION

LAG   COEF    T-VAL   +----+----+----+----+----0----+----+----+----+----+  LAG
 1  0.181     0.83    |                         |>>>>>>>>>>                |   1
 2 -0.095    -0.44    |                   <<<<<<|                          |   2
 3 -0.004    -0.02    |                         |                          |   3
 4 -0.162    -0.74    |               <<<<<<<<<<|                          |   4
 5 -0.378    -1.73    |<<<<<<<<<<<<<<<<<<<<<<<<<|                          |   5
                      +----+----+----+----+----0----+----+----+----+----+  LAG

               ACTUAL versus FITTED VALUES and RESIDUALS

SEQ    Actual   Fitted  min=41.900                   71.873=max  Error
        (*)      (+)    +----+----+----+----+----+----+----+----+              +--+--+
 1     41.90    42.22   |@                                       |   -0.324    |  <:  |
 2     45.00    46.25   |    *+                                  |   -1.250    |<<<:  |
 3     49.20    50.77   |         *  +                           |   -1.566    |<<<:  |
 4     50.60    51.09   |            @                           |   -0.494    |  <:  |
 5     52.60    52.59   |              @                         |    0.008    |   :> |
 6     55.10    54.23   |                +*                      |    0.869    |   :>>|
 7     56.20    54.86   |                 +  *                   |    1.338    |   :>>>
 8     57.30    56.25   |                   +*                   |    1.055    |   :>>|
 9     57.80    58.39   |                     *+                 |   -0.589    | <<:  |
10     55.00    54.72   |                 @                      |    0.282    |   :> |
11     50.90    51.13   |            @                           |   -0.230    |  <:  |
12     45.60    45.92   |     *+                                 |   -0.322    |  <:  |
13     46.50    46.18   |      @                                 |    0.322    |   :> |
14     48.70    48.76   |         @                              |   -0.058    |  <:  |
15     51.30    51.33   |            @                           |   -0.035    |  <:  |
16     57.70    56.08   |                   +*                   |    1.616    |   :>>>
17     58.70    59.14   |                      *+                |   -0.436    |  <:  |
18     57.50    57.29   |                     @                  |    0.210    |   :> |
19     61.60    60.61   |                         +  *           |    0.989    |   :>>|
20     65.00    64.21   |                              +*        |    0.785    |   :>>|
21     69.70    71.87   |                                     * +|   -2.173    <<<<:  |
        (*)      (+)    +----+----+----+----+----+----+----+----+              +--+--+

           LIST OF OUTLIERS - STUDENTIZED RESIDUALS GREATER THAN 1.5

 SEQ    Actual  Forecast   Residual  Std. Err  Studentized Probability
                                                Residuals     DF=17
 ----   ------  --------   --------  --------  ----------  ----------
    3    49.20     50.77     -1.566     0.967      -1.619       0.124
   16    57.70     56.08      1.616     0.958       1.687       0.110
   21    69.70     71.87     -2.173     0.780      -2.787       0.013
 ----   ------  --------   --------  --------  ----------  ----------

                    Covariance-Variance Matrix

                Profits        P-1        W+W'     Constant

  1  Profits     0.0083    -0.0053    -0.0013      0.0005
  2      P-1    -0.0053     0.0082    -0.0007     -0.0177
  3      W+W'   -0.0013    -0.0007     0.0016     -0.0329
  4 Constant     0.0005    -0.0177    -0.0329      1.6970

[Note: this sample output shows some of the options for regression diagnosis.]
```

MING TELECOMPUTING INC.

Telecommunications and Statistics for Microcomputers

23 Oak Meadow Road, P.O. Box 101, Lincoln Center, MA 01773, U.S.A. (617) 259-0391

Please send: ( ) ySTAT Package for $395. (Massachusetts residents: please add 5% sales tax.)
( ) ySTAT Trial Disk and information/sample output for $5.
( ) ySTAT information/sample output.

System: ( ) IBM PC. ( ) IBM XT. ( ) IBM AT. ( ) other ____________
( ) with 8087 or 80287 math co-processor. Memory ______ K

Payment: ( ) a check is enclosed.
( ) Visa. ( ) MasterCard. Credit Card #____________ Expiration Date: ___/___
( ) a university or governmental agency purchase order enclosed.

Name ____________ Signature (if charged) ____________

Address ____________

____________ Telephone: ( ) ____________

IBM is a registered trademark of International Business Machines Corp., 1-2-3 of Lotus, SAS of SAS Institute

Volumes in
ECONOMIC THEORY, ECONOMETRICS, AND MATHEMATICAL ECONOMICS

Edited by Karl Shell

FORECASTING ECONOMIC TIME SERIES, *Second Edition*

C.W.J. Granger and Paul Newbold

The very popular and widely-cited first edition of this book has now been updated to reflect the many recent developments in time series analysis and forecasting theory and practice, particularly as applied to economics. The major changes in this second edition are in the theory and application of multiple-series modeling procedures. The chapter on multiple-series modeling has been expanded to two chapters, and the final chapter provides a survey of recent and advanced topics that are currently of interest.
October 1986, 368 pp. $24.95 Paperback/ISBN: 0-12-295184-0
$49.50 Casebound/ISBN: 0-12-295183-2

MACROECONOMICS: AN INTRODUCTION TO THE NON-WALRASIAN APPROACH

Jean-Pascal Benassy

This book gives a clear exposition of non-Walrasian microeconomic foundations and then applies them to a large array of traditional macroeconomic topics by constructing simple, workable, and policy-oriented macroeconomic models. Applications include involuntary unemployment, inflation, expectations, cycles, and international economics.
June 1986, 272 pp. $26.95 Paperback/ISBN: 0-12-086426-6
$49.50 Casebound/ISBN: 0-12-086425-8

FORTHCOMING:

HOUSEHOLD AND ECONOMY: WELFARE ECONOMICS OF ENDOGENOUS FERTILITY

Marc Nerlove, Assaf Razin, and Efraim Sadka

This book integrates the theory of population into the main body of welfare economics. It is concerned with the normative implications of the so-called "new home economics"—the economics of family decisions about numbers of children, investments in their upbringing, labor force participation, and household production.
December 1986, 168 pp. $29.95 tentative. Casebound/ISBN: 0-12-515752-5

NONLINEAR ECONOMIC DYNAMICS

Edited by Jean-Michel Grandmont

The proceedings of a conference held in June 1985. Reprinted from the October 1986 issue of the *Journal of Economic Theory.*
December 1986, 165 pp. $19.95 tentative.

ACADEMIC PRESS
Harcourt Brace Jovanovich, Publishers
Orlando, FL 32887-0017

40126

New From Cambridge...

The Economics of Imperfect Competition

Melvin L. Greenhut, Chao-Shun Hung, George Norman

This book takes a new "spatial" approach to traditional price theory and to the analysis of imperfect competition. It has been increasingly recognized that the perfectly competitive paradigm is inappropriate to the explanation of many "real life" markets characterized by a significant separation between producers and consumers. The spatial perspective implies a natural separation of markets, but is a useful analogy for non-spatial issues. The analysis and results are shown to have significant policy implications.
Cloth about $37.50, Paperback about $16.95

Growth, Accumulation, and Unproductive Activity

An Analysis of Post-war U.S. Economy

Edward N. Wolff

This book documents the shift toward unproductive activity since World War II and its relation to the economic surplus, capital accumulation, and economic growth. The shift, which results in a reduction in new capital formation, in productivity growth, and in an erosion in the rate of growth in per capita living standards, is seen to be rooted in the logic of advanced capitalism.
About $37.50

Foreign Aid, War, and Economic Development

South Vietnam, 1955-1975

Douglas C. Dacy

Professor Dacy traces the economic history of South Vietnam from 1955 to 1975. He analyzes the impact of foreign and economic aid to the country, focusing on national income, as well as on a number of development indicators. He discusses the goals of United States economic aid, and measures the net resources transferred.
$39.50

Multinational Enterprise in Historical Perspective

Editors: Alice Teichova, Maurice Levy Leboyer, Helga Nussbaum

Based on extensive empirical research, this book provides an important overview of the origins and expansion, as well as the concept of, the "multinational" enterprise. Topics addressed include growth – both industrial and financial, market linkages with cartels, the relationship to government, and the international hierarchical and sexual division of labor. The work concludes by placing the multinational in historical perspective.
$49.50

Statistical Foundations of Econometric Modelling

Aris Spanos

This book provides an introduction to microeconomics by presenting its statistical foundations: probability theory and statistical inference. By emphasizing the underlying concepts, Professor Spanos promotes both a formal grounding and an intuitive understanding of these areas so essential to a proper understanding of economic theory.
$34.50

The Standard of Living

Tanner Lectures, Clare Hall, Cambridge 1985

Amartya Sen

Geoffrey Hawthorn, Editor

Amartya Sen argues that "the standard of living" has been poorly understood and defined. It is not just a function of opulence, and cannot be seen as utility, but rather as the "capabilities" offered in states of affairs. In reply, Bernard Williams, Ravi Kanbur, John Muellbauer, and Keith Hart each explore the concepts and implications of Sen's theory. Sen concludes with replies to these comments.
About $24.95

Selected Essays on Economic Planning

Michal Kalecki

Jan Toporowski, Editor and Translator

This work brings together the series of papers Kalecki wrote between 1940 and the 1960s on economic planning. They contain the germ of his theory of growth in a socialist society, as well as a startlingly original analysis of economic planning.

In his introduction, Jan Toporowski argues that Kalecki's work on growth is incomplete without the analysis presented in these papers, and has consequently often been misrepresented.
About $39.50

Essays on Economic Decisions Under Uncertainty

J. Drèze

This work contains selections of the author's contribution to the economic theory of decision under uncertainty, including an important new essay on "Decision theory under moral hazard and state-dependent preferences." Subsequent chapters treat decisions by consumer, producer, worker, and non-profit organizations.
About $39.50

Money, Wealth and Expenditure

Integrated Modelling of Consumption and Portfolio Behaviour

Dorian Owen

This book develops and applies a modelling framework, drawn from the Yale approach to financial model building, designed to analyze consumers' expenditure and portfolio behavior.
$34.50

Unemployment, Search and Labour Supply

Editors: Richard Blundell and Ian Walker

This book brings together the recent work of the foremost researchers in the field of labor economics. It analyses the labor market behavior of agents, particularly with regard to unemployment, job search, and labor market supply, and considers the impact of changes in economic and demographic factors.
$44.50

At your bookstore or call our Customer Service at 1-800-431-1580.
In NY and Canada 914-235-0300. MasterCard and Visa accepted.

Cambridge University Press

32 East 57th Street, NY, NY 10022

AMERICAN ECONOMIC ASSOCIATION

1987 ANNUAL MEMBERSHIP RATES

Membership includes:

—a subscription to both *The American Economic Review* (quarterly) plus *Papers and Proceedings* and the *Journal of Economic Literature* (quarterly).

● Regular members with annual incomes of $30,000 or less	$38.50
● Regular members with annual incomes above $30,000 but no more than $40,000	$46.20
● Regular members with annual incomes above $40,000	$53.90
● Junior members (available to registered students for three years only). Student status must be certified by your major professor or school registrar	$19.25
● In Countries other than the U.S.A., Add $12.00 to cover postage.	
● Family members (persons living at the same address as a regular member, additional memberships without subscription to the publications of the Association)	$7.70

Please begin my issues with:

☐ **March** ☐ **June** *(Includes Papers and Proceedings)* ☐ **September** ☐ **December**

First Name and Initial ____________ **Last Name** ____________ **Suffix** ____________

Address Line 1 ____________

Address Line 2 ____________

City ____________

State or Country ____________ **Zip/Postal Code** ____________

MAJOR FIELDS (TWO ONLY)
LIST FIELDS WITH WHICH YOU CURRENTLY IDENTIFY. SELECT FIELD CODE FROM *JEL*, "Classification System for Books."

☐ ☐

Please type or print information above. Please pay with a check or money order payable in United States Dollars. Canadian and foreign payments must be in the form of a draft or check drawn on a United States bank payable in United States Dollars. Please note: It is the policy of the Association, not to refund membership payments.

Endorsed by (AEA member) ____________

Below for Junior Members Only

I certify that the person named above is enrolled as a student at ____________

____

____
Authorized Signature

PLEASE SEND WITH PAYMENT TO:

AMERICAN ECONOMIC ASSOCIATION
1313 21ST AVENUE SOUTH, SUITE 809
NASHVILLE, TENNESSEE 37212-2786
U.S.A.

REMINDER PREPAYMENT OF 1987 MEMBERSHIP DUES MAY LOWER YOUR U.S. PERSONAL INCOME TAXES

PREPAYMENT OF 1987 DUES WILL BE ACCEPTED WITHOUT YOUR RENEWAL NOTICE.

Payment of 1987 annual membership dues in 1986 may be to your tax advantage because of the reform of the Tax Code. Your prepayment of 1987 dues will be accepted without your renewal notice. When paying, please include your account number, if known, and the address which is currently on file for you. You may include an address change if applicable.

Please refer to the 1987 Annual Membership Rate schedule on the preceding page.

When the subject of economics comes up in a conversation...

With the most extensive software package available...

Economics, 3/e

Ralph T. Byrns / Gerald W. Stone, Jr.
Metropolitan State College and The University of Colorado at Denver

Available Now. *©1987, 780 pages, hardbound. Also available in separate macro/micro softbound editions.*

- New MACROSIM II and RAIDERS (computer simulations), each in student and instructor versions. Student version includes 120-page booklet that drills core text ideas leading into use of the simulation.
- The most "reader-friendly" writing style of any introductory text on the market.
- Timely boxed examples throughout the text on such topics as liability insurance crisis, Congressional tax bill of 1986, the New Poor, changing rules in antitrust, and Mexican debt and capital flight.
- Early introduction of aggregate supply and aggregate demand (Ch. 6) for repeated use as a macro tool, plus solid treatment of Keynesian economics and business cycles.
- Unique micro section includes legal cases (with their economic underpinnings) at the end of many chapters.

Modern Managerial Economics

Martin T. Farris / Stephen K. Happel
Arizona State University

This text stresses neoclassical theory and strategic use of managerial decision making. Strong coverage of real-world pricing and managing uncertainty. Separate chapters on location decisions and nonprofit firms. Optional math and calculus sections. **Available Now.** *©1987, 590 pages, hardbound.*

Microeconomics and Public Policy

William C. Apgar / H. James Brown
Harvard University

This unique text focuses on the economics of public policy and public sector decision making, as well as the core tools of neoclassical microeconomics. Two chapters on cost-benefit analysis are included. **Available Now.** *©1987, 410 pages, hardbound.*

Scott, Foresman Economics • Scott, Foresman Economics

So does Scott, Foresman.

American Economic History, 2/e

Jonathan Hughes, Northwestern University

This beautifully written best-seller offers a new chapter on the manufacturing industry and its evolution since World War II, as well as important new material on the postbellum Southern economy and the Great Depression. **Available Now.** *©1987, 632 pages, hardbound.*

Investment: Concepts, Analysis, and Strategy, 2/e

Robert C. Radcliffe, University of Florida

The most readable and analytic investments text offers worked-out problem sets, new portfolio theory chapters, and two unique new chapters on portfolio management. **Available Now.** *©1987, 760 pages, hardbound.*

Money, Financial Institutions, and Economic Activity

Bruce R. Dalgaard, University of Minnesota, Twin Cities

A unique industrial organization approach emphasizes the microeconomics of industry competition. Strong historical perspective on institutions and policy shows the evolving nature of the financial industry. Includes the latest information on financial futures, deregulation, and the microeconomics of interest rate determination. **Available February.** *©1987, 608 pages, hardbound.*

Options: Theory, Strategy, and Applications

Peter Ritchken, Case Western Reserve University

A comprehensive introduction to the management of portfolios using options in financial markets. Stock, index, debt, and foreign currency options are discussed, as well as forward and futures contracts and options on these instruments. **Available Spring 1987.** *©1987, 320 pages, softbound.*

Recently Published

Intermediate Microeconomics: Theory and Applications, 2/e

Heinz Kohler, Amherst College

The most applied price theory text has been revised: new I.O. work, carefully wrought calculus footnotes/appendices, new layout, and three microcomputer (IBM) disks of challenging problem sets. *©1986, 608 pages, hardbound.*

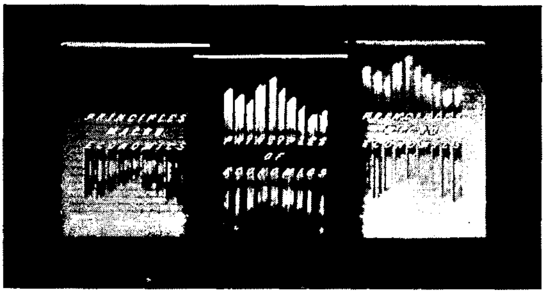
Intermediate Macroeconomics

Michael R. Darby, University of California, Los Angeles
Michael T. Melvin, Arizona State University

The only text to cover both new and traditional macro topics within the fully teachable length of less than 500 pages. Superb chapter on microfoundations; three chapters on dynamic adjustments between short and long run. *©1986, 455 pages, hardbound.*

Scott, Foresman Economics • Scott, Foresman Economics

Take a close look at the new best-selling principles text...



Principles of Economics, 2/e
Roy J. Ruffin / Paul Gregory
University of Houston

Available Now. *©1986, hardbound, with new Microcomputer Modules. Also available in separate macro/micro editions.*

- New, full chapter on interest rates
- Aggregate supply/aggregate demand model reworked and integrated within text
- Increased coverage of federal deficits, Volcker's monetary policies, and Reagan's new flat-tax plan
- New agent/principle theme introduced throughout I.O. chapters, plus new I.O. theories
- Nash equilibrium, moral hazard, and adverse selection problems
- 100 new boxed examples

Macroeconomics
David Colander, Middlebury College

This intermediate-level text approaches modern macroeconomics through its historical evolution: classical model, Keynesian model, and monetarist model, followed by New Classical and New Keynesian economics. Strong institutional and policy setting. *©1986, 545 pages, hardbound.*

Essentials of Economics
Paul Gregory / Roy J. Ruffin
University of Houston

The first one-term text to build its macro models from the micro foundations presented earlier in the text. Modern industrial organization, economics of information, integrated aggregate supply/demand, expectations theory, and contracting are covered. *©1986, 448 pages, softbound.*

For further information write
Meredith Hellestrae, Department SA—AER
1900 East Lake Avenue
Glenview, Illinois 60025

Scott, Foresman and Company

Scott, Foresman Economics • Scott, Foresman Economics

Leadership at the Fed

Donald F. Kettl

A colorful and fascinating history of one of Washington's most powerful agencies and of the men who enabled it to grow into the national and international force it is today. Making use of previously untapped sources, Kettl provides a unique look at the leaders of the Federal Reserve Board, from Marriner Eccles and William McChesney Martin to G. William Miller and Paul Volcker.

"**A well-organized, well-written, and important** contribution to the literature on Federal Reserve policy making." —Ambassador Arthur F. Burns $22.50

The First Resource

Wild Species in the North American Economy

Christine Prescott-Allen
and Robert Prescott-Allen

The first systematic analysis of the economic importance of wildlife to North America.

"The authors have approached a neglected, usually ignored topic with a thoroughness of detail supported by an incredible wealth of statistical material. **A masterly production.**" —Richard Evans Schultes Illus. $65.00

Sponsored by the World Wildlife Fund

Common Decency

Domestic Policies After Reagan

Alvin L. Schorr
with a chapter by James P. Comer

An incisive critique of America's social policies under the Reagan administration that offers clear, reasonable, and workable proposals for improving our policies in such areas as income distribution, housing, health, and education.

"**A masterful work of policy analysis.** Schorr brings penetrating insights and humanistic sympathies to bear on the central issues in social welfare policy." —Neil Gilbert $20.00

Yale University Press
Dept. 363
92A Yale Station
New Haven, CT 06520

In Whose Interest?

International Banking and American Foreign Policy

Benjamin J. Cohen

In the first book to investigate the potential for conflict between the interests of American banks and the goals of American foreign policy, a distinguished economist presents innovative and hard-hitting policy recommendations for improving this relationship and offers practical suggestions for dealing with the most serious bank-related issues on Washington's current foreign-policy agenda.

"Cohen is . . . one of the few international economists who is sufficiently sensitive to politics and the role of institutions to write a cogent, policy-related study on this subject. . . . **Valuable.**" —Robert O. Keohane
$19.95

A Council on Foreign Relations Book

Biotechnology: The University-Industrial Complex

Martin Kenney

"**A fascinating** study of the 'shattering of the ideology of pure science under the impact of economics.' " —Hilary D. Burton, *Library Journal*

"Here is a dispassionate discussion of biotechnology — its beginnings and where it may lead. Dr. Kenney's analysis of the interrelations between corporate and university forces is **thoughtful and comprehensive.** The book . . . should provide the basis for the socioeconomic debate that will inevitably intensify with time." —Liebe F. Cavalieri $23.95

Journal of Law, Economics, and Organization

Jerry Mashaw and Oliver Williamson, co-editors

One-year subscription (two issues) $28.00 institutions, $20.00 individuals. $14.00 per issue.

Uncompromising statistics

Mainframe statistical packages are mostly used out of habit. Microcomputer statistical packages are used just because they are available. It's time to pick a statistical environment because it suits *your* needs, whether it be on a micro or a mainframe. And it doesn't have to cost a lot.

Statistical Software Tools is an integrated, interactive environment for database management, data analysis, estimation and inference. Without compromises. Available now for MS-DOS, Unix and VAX/VMS computers.

Enter data interactively or transfer files directly from other popular programs. Full data manipulation capabilities and a convenient system file format.

Tabulate data. Compute descriptive statistics. Regressions and regression diagnostics without hassle. Character plotting for all devices. Full graphics if you have a graphics card.

Statistical procedures include ordered probit and logit, multinomial logit, maximum likelihood estimation for user-defined problems. Econometric estimation including simultaneous equations, Tobit models and random utility models.

Fully programmable. User defined functions and macros. Procedures including control loops. And much more.

Powerful computing for those who know how to use it. Yet easy to use with a simple command language. All clearly and thoroughly explained in a *User's Guide* and *Reference Manual.*

There is only one feature which is limited. The price. Only $100 for the IBM PC and compatibles. Requires 384K RAM. 8087 numeric coprocessor recommended. Order from Dubin/Rivers Research, 1510 Ontario Avenue, Pasadena, California 91103. Or call (818)577-8361.

sst

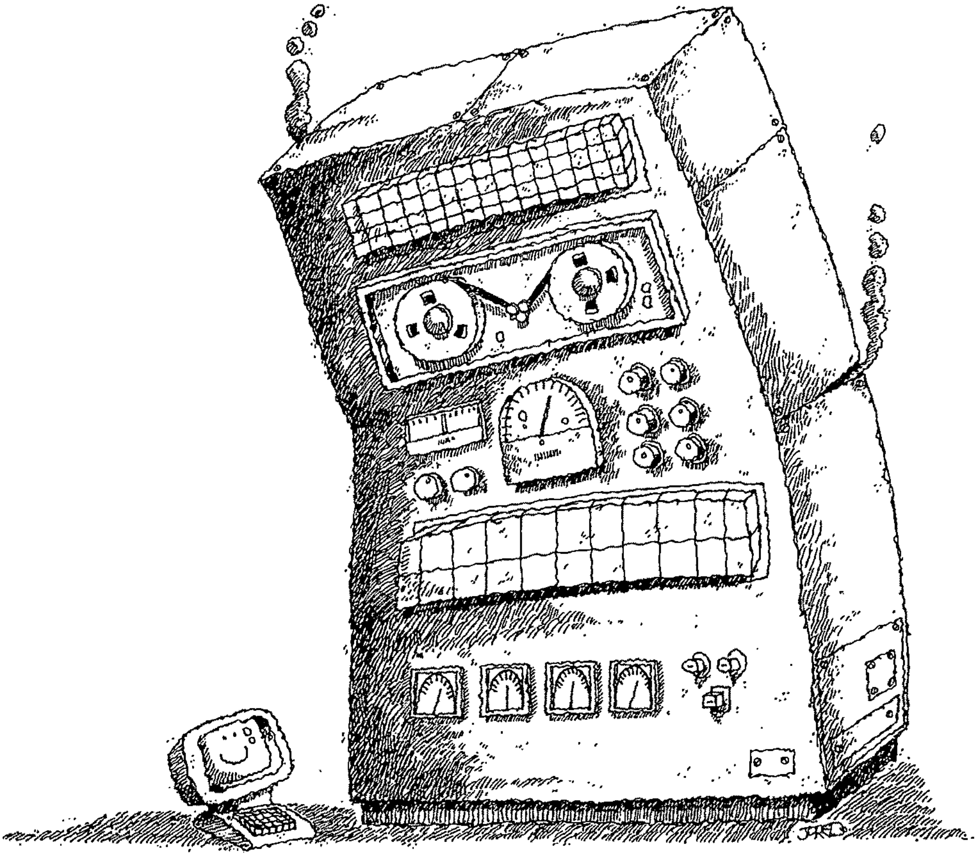
MS-DOS is a trademark of Microsoft Corp. Unix is a trademark of AT&T Bell Laboratories. VAX is a trademark of Digital Equipment Corp. IBM PC is a trademark of International Business Machines Corp.

FEDERAL RESERVE BANK OF NEW YORK

MAY 18 1914

Positions available for economists with strong publication record in macroeconomics, money, finance or international economics. Temporary and permanent appointments available. Salaries commensurate with experience and level of achievement. For more information contact:

Christine Cumming
Recruiting Officer
Federal Reserve Bank of New York, Rm. 922
New York, New York 10045
(212) 720-8258



Bigger does not always mean better.

The first computers were so big they filled entire rooms, and had to be shut down periodically because of overheating.

Things got better when things got smaller. And that's something to consider if you're looking for a publisher.

Now, we at South-Western Publishing would never suggest you select a publisher solely based on size. We would, however, suggest you select a publisher based on quickness of reaction to your needs, personalized service, and flexibility. Because your book isn't like everyone else's. So it shouldn't be treated that way.

So call us at (513) 527-6384. And remember that while some of the big, prodigious publishers may look impressive, they might not respond to your needs the way we will. The important thing about a publishing company is not how big it is, but how big it thinks.

SOUTH-WESTERN
COLLEGE DIVISION

Economics for Professional Economists . . .

from Praeger.

NOT SO FREE TO CHOOSE
The Political Economy of Milton Friedman and Ronald Reagan
Elton Rayack

Here is a critical and carefully documented study of Friedman's writings. Claiming that the influential economist's writings have exerted a powerful influence on the policies, ideology, and rhetoric of the Reagan administration, the author examines some 300 columns Friedman has written for *Newsweek* along with his best-selling books in order to determine the extent of his influence on the administration.
176 pages $35.85 (tent.) December 1986

AMERICA'S CHANGING ROLE IN THE WORLD SYSTEM
edited by **Terry Boswell** and **Albert Begesen**

This volume brings together 13 experts in world-systems analysis to examine the long-term impact of America's declining position in the world economy.
272 pages $35.85 (tent.) December 1986

TRADING TECHNOLOGY
Europe and Japan in the Middle East
Thomas L. Ilgen and T.J. Pempel

Here is an in-depth study of high technology trade between four major industrialized countries — Britain, France, West Germany, and Japan — and the oil producing states of the Middle East. Along with an insightful analysis of the general economic interaction between the industrialized nations and the Middle East, the book offers a thought-provoking chapter on implications for U.S. policy.
192 pages $35.85(tent.) December 1986

THE CARIBBEAN EXODUS
edited by **Barry B. Levine**

A timely look at the complexities of the Caribbean migration, its impact on the region and on the U.S.
cloth: 240 pages $32.85 (tent.) December 1986
paper edition: $12.85 (tent.) December 1986

Praeger Publishers
A Division of Greenwood Press, Inc.
P.O. Box 5007 • 88 Post Road West • Westport, CT 06881 • 203-226-3571

AMERICAN GREEN POWER

by Alain Revel and Christophe Riboud
translated by Edward W. Tanner

"Overflowing silos on the one side, starving multitudes on the other: such is the mad situation revealed in AMERICAN GREEN POWER, an essential work for the understanding of the next twenty years."—*L'EXPRESS*

Alain Revel and Christophe Riboud examine the world food balance and explore the sources and implications of U.S. "green power," an agricultural strength that no other country can seriously challenge.

$32.50 ***hardcover***
$14.95 ***paperback***

ECONOMIC VALUATION TECHNIQUES FOR THE ENVIRONMENT

A Case Study Workbook

edited by John A. Dixon and Maynard M. Hufschmidt

The authors set forth a methodology for assessing and valuing the environmental effects of development projects in *Environment, Natural Systems, and Development: An Economic Valuation Guide.* Now they present detailed case studies demonstrating applications of those techniques in a workbook designed to serve both as a companion volume and as an independent collection of instructive examples.

$25.00 ***hardcover***
$8.95 ***paperback***

ECONOMIC APPRAISAL OF TRANSPORT PROJECTS

A Manual with Case Studies

REVISED AND EXPANDED EDITION

by Hans A. Adler

This classic textbook, now updated, provides a practical introduction to the objectives and methodology of the economic appraisal of transport projects in developing countries. Fifteen case studies—treating road, bridge, railway, pipeline, port, and airport projects—exemplify the kinds of problems typically encountered and illustrate the concepts and techniques described.

A Publication of the World Bank

$27.50 ***hardcover***
$12.50 ***paperback***

THE
JOHNS HOPKINS
UNIVERSITY PRESS

701 West 40th Street, Suite 275, Baltimore, Maryland 21211

Oxford

World Economic Outlook 1986

A survey by the staff of the INTERNATIONAL MONETARY FUND

The *World Economic Outlook* is an important supplementary book for courses in international economics and trade. Packed with the latest statistical information and insightful discussions of the major issues affecting today's economy, it supplies the data that will make theoretical courses in international economics real to students.
1986 268 pp.; charts and graphs paper $11.00*

Microeconomic Principles

FRANK A. COWELL, *The London School of Economics and Political Science*

This comprehensive textbook offers a thorough coverage of the basic principles that underlie standard modern theoretical and applied economics. The book includes exercises with answers to help students understand and apply the analytic techniques presented.
1986 360 pp. cloth $24.95

Input-Output Economics

Second Edition

WASSILY LEONTIEF, *Former Director, Institute for Economic Analysis, New York University*

In 1973 Wassily Leontief was awarded the Nobel Prize for his model of input-output economics. This collection of papers—the only comprehensive introduction to the model he has written—is now thoroughly revised to include thirteen new essays reflecting the development and present state of the field.
1986 320 pp.; 30 line illus. paper $16.95* cloth $29.95

Game Theory with Applications to Economics

JAMES W. FRIEDMAN, *Virginia Polytechnic Institute and State University*

Drawing on examples from current economic literature and politics, this is the first book on game theory at an introductory, but not elementary, level. The author covers those topics in game theory that are of great actual or potential use in economics, such as noncooperative games, two-person cooperative games, and cooperative games with and without side payments.
1986 292 pp.; 41 figs. cloth $24.95

Now available in paperback

History of Economic Analysis

The late JOSEPH SCHUMPETER

A complete history of man's theoretical efforts from the time of ancient Greece through the present to understand economic phenomena, *History of Economic Analysis* is an important contribution to the history of ideas as well as to economics.
1986 1,288 pp. paper $19.95*

Choosing the Right Pond

Human Behavior and the Quest for Status

ROBERT H. FRANK, *Cornell University*

"A landmark event."—*The Boston Globe*. "An exceptional book."—*Los Angeles Times*. This challenging and provocative book offers an alternative to the prevailing view of human beings as economic automatons, demonstrating instead how individual desires—notably the quest for status—profoundly affect the marketplace.
1985 (paper December 1986) 244 pp. paper $9.95* cloth $22.95

From the series *Europe and the International Order*

Reagan, Thatcher, and the Politics of Decline

JOEL KRIEGER, *Wellesley College*

Examining the vital and confident conservatism that has come to replace the more consensual basis of political life that typified the postwar years in both the United States and Britain, Joel Krieger argues that the emergence of a "New Right" in both countries and the consolidation of these highly ideological governments are responses to long-term decline in economic competitiveness and geopolitical influence, focused in an acute way by the global recession.
1986 320 pp. cloth $29.95 paper $11.95

Governing the Economy

The Politics of State Intervention in Britain and France

PETER A. HALL, *Harvard University*

For over one hundred years, the British economy has been in decline relative to other industrialized countries. France, on the other hand, experienced an economic miracle in the postwar period. Peter Hall examines and compares the economic policy and state intervention instituted by both countries, developing a new approach to state-society relations that emphasizes the crucial role of institutional structures.
1986 340 pp. cloth $36.00 paper $14.95

The Politics of Neocorporatism in France

Farmers, the State, and Agricultural Policymaking in the Fifth Republic

JOHN KEELER, *University of Washington*

This penetrating study focuses on the emergence of the most powerful agricultural union (FNSEA) as a force in French politics, the trend toward corporatization of the agricultural sector, and the governmental efforts at reform and decorporatization in Mitterrand's France.
1986 320 pp. $49.95

The Welfare State East and West

Edited by RICHARD ROSE, *University of Strathclyde*, and REI SHIRATORI, *University of Essex* and *Dokkyo University*

This unique, comparative examination of the different ways in which modern democratic societies provide welfare today focuses on the welfare strategies and experiences of the U.S., Japan, Great Britain, Scandinavia, Germany, and Israel, discussing each society's ideas about the best way for the state, the market, and the household to contribute to the welfare mix.
1986 320 pp. $29.95

The Collected Works of Walter Bagehot

Volumes XII and XIII: Letters
Volumes XIV and XV: Miscellany

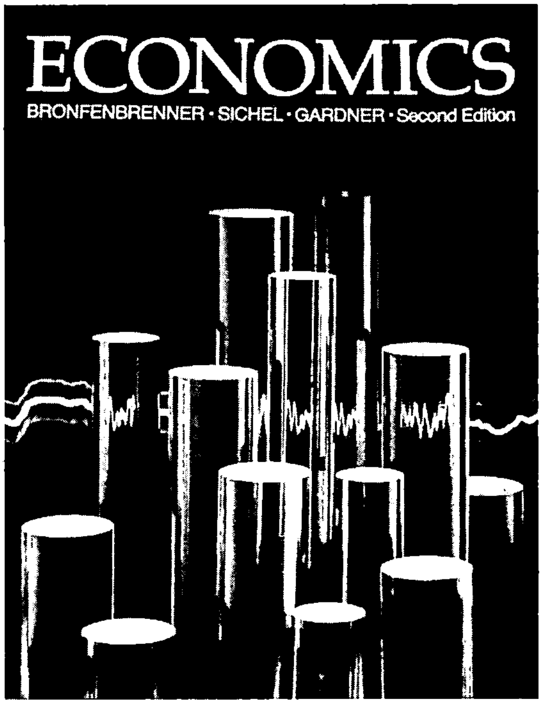
Edited by NORMAN ST. JOHN-STEVAS

Walter Bagehot (1828-1877), banker, economist, political thinker and commentator, critic and man of letters, was Victorian England's most versatile genius. The publication of his collected works is now complete with the addition of these four volumes. Volumes XII and XIII are devoted to the letters, half of which are previously unpublished. Volumes XIV and XV bring together articles and essays by Bagehot and writings about him
1986 Vols. XII and XIII: 736 pp.; 2 pp. plates $135.00 boxed set
Vols. XIV and XV: 932 pp.; 16 pp. plates $135.00 boxed set
complete set $750.00

Prices and publication dates are subject to change.
To order, send check or money order to: Humanities & Social Sciences Marketing Dept.
**To request an examination copy of books marked with an asterisk, write on school letterhead giving full course information, including course name, level, expected enrollment, and your decision deadline, to: College Humanities & Social Sciences Marketing Dept.*

Oxford University Press
200 Madison Avenue · New York, NY 10016

THE RATIONAL CHOICE



The first edition of Bronfenbrenner/Sichel/Gardner—the rational choice—brought the economics of the '80s into the mainstream of the principles course. Now, in its second edition, *Economics* provides an even stronger presentation of aggregate demand and aggregate supply, increases emphasis on applied microeconomics, and discusses current economic developments.

ECONOMICS, Second Edition
Martin Bronfenbrenner, Aoyama Gakuin University, Japan
Werner Sichel and **Wayland Gardner**
both of Western Michigan University
Complete hardcover edition: About 847 pages
Two-Volume paperback edition:
MACROECONOMICS • About 451 pages
MICROECONOMICS • About 572 pages
Complete support package available.
December 1986.

For adoption consideration, request an examination package from your regional Houghton Mifflin office.

Houghton Mifflin Company
13400 Midway Rd., Dallas, TX 75244-5165
1900 S. Batavia Ave., Geneva, IL 60134
989 Lenox Dr., Lawrenceville, NJ 08648
777 California Ave., Palo Alto, CA 94304

IS THE NATIONAL CHOICE.

If you're using a timesharing company for forecasting, you're paying a small fortune.

But you don't have to. Not with the Economica family of economic software products and your microcomputer.

Our economic software and models have been developed by leading experts, including the renowned Ray Fair. All models are fully proven in the marketplace. And you'll save thousands by using them.

You'll also save time. That's because you'll see the results of your modeling efforts immediately. Plus, you'll get great flexibility, there are no constraints. You can run alternate scenarios as often and whenever you wish. And accuracy, based on your input, is exceptional.

No matter what economic indicators you use for forecasting, we've got a solution. At the core of our family of economic software is *FAIRMODEL*, a quarterly model of the U.S. economy. Then there's *ESP® – The Econometric Software Package*,™ a tool for model building and economic forecasting. *FINANCIAL MODEL*, a money and interest rate model with 21 variables. *INDUSTRY MODEL*, it gives quarterly forecasts of industrial output for 90 sectors of the U.S. economy. *WORLD FAIRMODEL*, a quarterly model forecasting key macro variables for 63 countries with bilateral trade flows. And Citibase-Mini, economic data on diskettes.

At Economica we're totally dedicated to economic software. It's all we do. Our products are fully documented, easy to learn, and run on your own MS-DOS based microcomputer. No matter what other packages you've been using, learn how you can test ours. There are none better, or easier on your budget. Mail the coupon or call (617) 661-3260.

ECONOMICA
inc

Tell me how I can test Economica software. Please send me more information on ☐ *FAIRMODEL*, ☐ *ESP/The Econometric Software Package*, ☐ *FINANCIAL MODEL*, ☐ *INDUSTRY MODEL*, ☐ *WORLD FAIRMODEL*, ☐ *Citibase-Mini*, ☐ The entire Economica family of economic software.

Name____________________

Title____________________

Company____________________

Address____________________

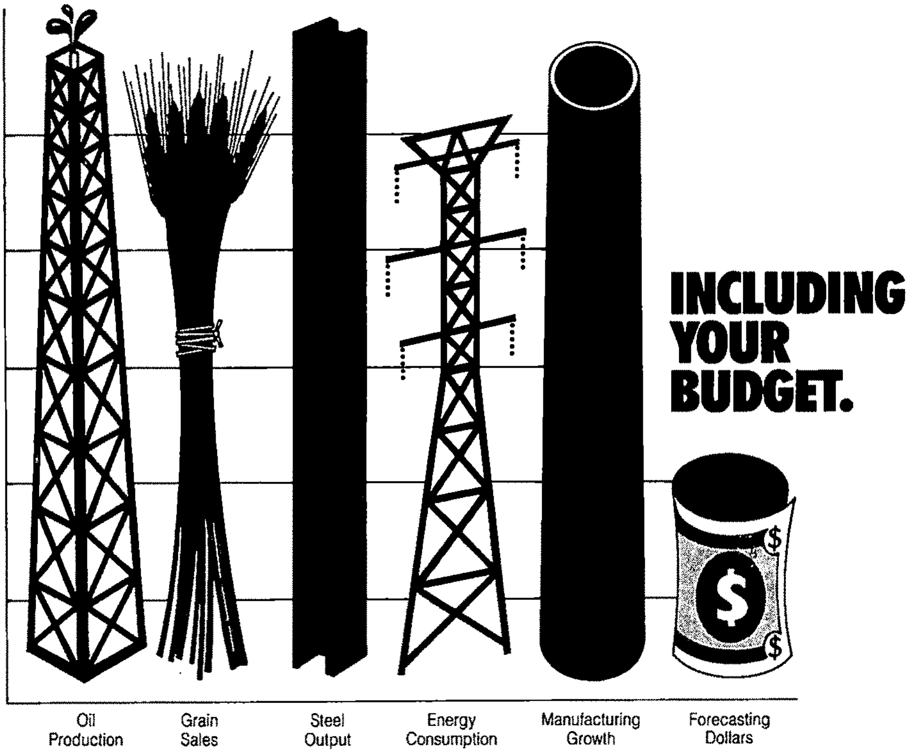
City__________ State_____ Zip________

Telephone________________ Ext.______

AER 12/86

ECONOMICA, 2067 Massachusetts Avenue, Cambridge, MA 02140, TELEX via WUI 6502773397

FINALLY, THERE ARE PC BASED FORECASTING TOOLS THAT RESPOND TO EVERY ECONOMIC INDICATOR.



American Economic Association/Federal Reserve System
Minority Graduate Fellowships
in Economics

The American Economic Association and the Federal Reserve System are pleased to announce their joint sponsorship of graduate fellowships to minority Ph.D. students who have completed their comprehensive examinations and, if applicable, their field examinations, and are about to begin their dissertation research. Awards will be based on academic performance.

Applicants must be U.S. citizens who are Black, Hispanic or Native American and are enrolled in an accredited graduate program in Economics in the United States. Preference will be given to applicants whose area of concentration is of special interest to the Federal Reserve System (e.g., financial markets and monetary policy, nonfinancial macroeconomics, forecasting, banking markets and financial structure, regional studies, the external sector of the U.S. economy, the economies of other countries, foreign exchange markets, and international banking and financial markets).

A stipend of $600 per month for the academic year and tuition relief by the institution nominating the student and verifying successful participation in its graduate program in Economics are provided. Recipients will be assigned an adviser from the Federal Reserve System and be given the opportunity to work for one summer at the Board or a Federal Reserve Bank.

Applications are due February 1, 1987
Awards will be announced early in April 1987

For further information and application materials write:

Barbara Sears, Registrar
AEA/FRS Minority Fellowship Program
Department of Economics
College of Business and Public Administration
University of Arizona
Tucson, AZ 85721
(602) 621-3272

New from the Consortium of Social Science Associations....

The first comprehensive

Guide to Federal Funding for Social Scientists

Prepared by the Consortium of Social Science Associations (COSSA), a Washington advocacy group serving the major professional societies in the social and behavioral sciences. Susan D. Quarles, editor.

The federal government is a major supporter of research in the social and behavioral sciences, but until now, no single, multidisciplinary directory has been available to guide researchers through the complexities of government funding in these fields.

COSSA's inclusive new *Guide to Federal Funding* describes over 300 federal programs in impressive detail, including funding priorities, application guidelines, and examples of funded research. Introductory essays describe the organization of social science funding and offer inside views of federal funding practices and contract research.

For anyone who needs to know the ins-and-outs of government funding in the social sciences and related fields, COSSA's *Guide* will be an essential new resource.

Published by the Russell Sage Foundation

Available in June 512 pages ISBN 0-87154-699-X Paperback

Mail orders to:

Consortium of Social Science
Associations
1200 Seventeenth St., NW, Suite 520
Washington, D.C. 20036

Orders will be filled when books are available, in June. Please allow 3-4 weeks for delivery.

Please send_____copies of COSSA's *Guide to Federal Funding for Social Scientists* (81-0699X) at the following price:

☐ $24.95 (for libraries and institutions)
☐ **$14.95 (special price for members of AEA)**
☐ $19.95 (for other individuals)

Enclosed is a check, money order, or purchase order in the amount of $_________. Publisher pays postage on prepaid orders; New York residents, please add sales tax.

____
Name (please print)

____
Address

____

A CLASS ACT

A Basic Framework for Economics ◁ NEW FOR '87
Leftwich
Gay *Third Edition*

Macroeconomics ◁ NEW FOR '87
Ekelund
DeLorme *Second Edition*

Economics of Social Issues
Sharp
Leftwich *Seventh Edition*

Managerial Economics
Southwick

Managerial Economics: Analysis and Cases
Coyne *Fifth Edition*

Readings In Managerial Economics
Coyne *Fourth Edition*

Intermediate Microeconomics
Solberg

Financial Institutions and Markets in a Changing World ◁ NEW FOR '87
Fraser
Rose *Third Edition*

Money and Capital Markets
Rose *Second Edition*

Financial Institutions
Rose
Fraser *Second Edition*

Money and Banking: Contemporary Practices, Policies and Issues
Black
Daniel *Second Edition*

Money, Financial Institutions and the Economy
Harrison

Public Finance: A Normative Theory
Tresch

Labor Relations: Development, Structure, Process
Fossum *Third Edition*

The Collective Bargaining Process: Readings and Analysis
Baderschneider
Block
Fossum

NEW FOR '87 ▷ **Statistics for Business and Economics**
Van Matre
Gilbreath *Third Edition*

Business and Economics Statistics
Plane
Oppermann *Third Edition*

Fundamentals of Management Science
Turban
Meredith *Third Edition*

Essentials of Management Science
Meredith
Turban

Cases and Readings in Management Science
Turban
Loomba *Second Edition*

BUSINESS PUBLICATIONS, INC.
1700 Alma, Suite 390, Plano, Tx 75075

BROOKINGS

The Comparable Worth Controversy
Henry J. Aaron and Cameran Lougy

This concise study reviews the conditions that have sparked the intense debate over the concept of comparable worth. The authors unravel the implications of comparable worth for employers in public and private sectors, for labor union agendas and employer-employee negotiations, and for the administrative and judicial burdens of the nation's courts.

1986/57 pp./paper $7.95

Saving Free Trade:
A Pragmatic Approach
Robert Z. Lawrence and Robert E. Litan

This volume provides a detailed but nontechnical introduction to the complex implications of amending trade policy and shrewd, innovative proposals for improving America's ability to adapt to rapid changes in the world markets.

1986/132 pp./paper $8.95/cloth $22.95

Other Times, Other Places:
Macroeconomic Lessons from U.S. and European History
Charles L. Schultze

These penetrating lectures illuminate three areas of economic controversy with insights gained from comparing and contrasting economic behavior in different eras and different countries.

1986/88 pp./paper $7.95/cloth $15.95

French Industrial Policy
William James Adams and
Christian Stoffaës, Editors

This volume provides a contentious, complex, balanced, and lively view of how government intervened in French industry and describes the implications of the French experience for both France and the United States. The contributors include members of the French government, civil service, and business community, and American observers.

1986/228 pp./paper $10.95/
cloth $28.95

The Swedish Economy
Barry P. Bosworth and
Alice M. Rivlin, Editors

Focusing on how high taxes and welfare programs affect economic efficiency and incentives, this comprehensive volume evaluates changes in the Swedish economy, from decades of rapid growth to severe problems during the 1970s.

1987/c. 350 pp./paper $12.95/
cloth $32.95

The Brookings Institution
1775 Massachusetts Avenue, N.W.
Washington, D.C. 20036
(202) 797-6258

BROOKINGS

Large Scale Modeling at IUI

TWO MODELS OF AN OPEN ECONOMY

by Lars Bergman, Leif Jansson, Tomas Nordström and Bengt-Christer Ysander (ed.)

Large scale, economy-wide econometric models are essential tools for addressing and analyzing many problems of economic policy. This book documents two such models. The first – ISAC – is a disequilibrium model, developed at IUI. The second – ELIAS – is an equilibrium model, developed at the Stockholm School of Economics. These two models represent different schools of thought in model-building. Both are designed, however, to capture the mechanisms through which disturbances in world markets induce structural adjustment in the Swedish economy.

THE FIRM AND FINACIAL MARKETS IN THE SWEDISH MICRO-TO-MACRO MODEL

— Theory, Model and Verification

by Gunnar Eliasson

The main ideas behind dynamic micro-macro theory are presented together with a detailed account of the model – called MOSES. The firm is emphasized as a financial decision unit that takes independent hiring, production, investment etc. decisions. Some 70 percent of Swedish manufacturing industry is represented at division levels. Aggregation up to national accounts level is endogenous through dynamic markets. Manufacturing is in turn integrated in a dynamic 11 sector economy-wide model with complete demand feedback. A summary of calibration methods, database work and empirical applications is also included.

IUI PUBLISHING

Industriens Utredningsinstitut
Grevgatan 34, S-114 53 STOCKHOLM, Sweden